

Pontificia Universidad Católica del Ecuador

Facultad de Economía

Disertación previa a la obtención del título de Economista

Caracterización de la demanda de servicios de salud en el Ecuador. Una aplicación empírica.

Diego David Villacreses Zúñiga

ddvillacresesz@gmail.com

Director: Mtr. Juan Pablo Erráez

jperraez@puce.edu.ec

Quito, mayo de 2015

Resumen

La presente disertación es un análisis empírico de la demanda de salud en el Ecuador en base a la Encuesta Nacional de Salud y Nutrición 2011-2013. Se explora la teoría desarrollada por Grossman (1972) y se ejecuta la misma aplicación empírica que el autor para el Ecuador, encontrando los mismos resultados. Además se amplía el análisis al utilizar modelos de conteo y un conjunto más amplio de variables independientes. Se encuentra que variables como el acceso a agua potable y la presencia de un cuarto exclusivo para cocinar son de importancia. Además se demuestra la importancia de la actividad física para la salud y considerables diferencias territoriales entre las regiones Costa y Sierra.

Palabras Clave: modelo de Grossman, demanda de salud, aplicación empírica, técnicas econométricas.

A todos los que aportaron con algo de felicidad a mi vida

A mis padres por su amor incondicional

Caracterización de la demanda de servicios de salud en el Ecuador. Una aplicación empírica

Introducción	7
Metodología de Investigación	10
Pregunta general	10
Preguntas específicas.....	10
Objetivo General.....	10
Objetivo Específico	11
Fundamentación Teórica	12
Bienes Públicos	12
Modelos de Predicción.....	12
Modelos de Variables Binarias.....	14
Modelos de conteo	17
El problema de la estimación causal.....	19
Análisis Causal	24
Salud y economía	26
Demanda de Salud.....	28
Planteamiento Empírico.....	30
Modelo de Grossman.....	30
Otros determinantes de la Demanda de Salud.....	42
Modelos de Conteo para análisis de intensidad de enfermedades y demanda de salud.....	55
Modelos de Conteo para análisis de intensidad de enfermedades y demanda de salud con un enfoque territorial.....	60
Conclusiones.....	64
Recomendaciones.....	65
Bibliografía	66
Anexos	69
Anexo No. 1	69
Anexo No. 2.....	70
Matrices de confusión, curva de sensibilidad y curva ROC para modelo de probabilidad de enfermarse.	70
Matrices de confusión, curva de sensibilidad y curva ROC para modelo de demanda de servicios de salud.....	72

Matrices de confusión, curva de sensibilidad y curva ROC para modelo de probabilidad de enfermarse al incluir variable de actividad física.....	74
Matrices de confusión, curva de sensibilidad y curva ROC para modelo de demanda de servicios de salud al incluir variable de actividad física.....	76
Anexo 3	78
Sintaxis para la obtención de estimaciones	78

Índice de ilustraciones

Ilustración No. 1	31
Ilustración No. 2	32
Ilustración No. 3	46
Ilustración No. 4	47
Ilustración No. 5	51
Ilustración No. 6	54
Ilustración No. 7	70
Ilustración No. 8	71
Ilustración No. 9	72
Ilustración No. 10	73
Ilustración No. 11	74
Ilustración No. 12	75
Ilustración No. 13	76
Ilustración No. 14	77

Índice de tablas

Tabla No. 1.....	7
Tabla No. 2.....	8
Tabla No. 3.....	17
Tabla No. 4.....	36
Tabla No. 5.....	37
Tabla No. 6.....	40
Tabla No. 7.....	41
Tabla No. 8.....	43
Tabla No. 9.....	46
Tabla No. 10.....	48
Tabla No. 11.....	50
Tabla No. 12.....	52
Tabla No. 13.....	54
Tabla No. 14.....	57

Tabla No. 15.....59

Tabla No. 16.....61

Tabla No. 17.....63

Tabla No. 18.....70

Tabla No. 19.....71

Tabla No. 20.....72

Tabla No. 21.....73

Tabla No. 22.....74

Tabla No. 23.....75

Tabla No. 24.....76

Tabla No. 25.....77

Introducción

La salud de las personas afecta significativamente al desarrollo de un país, ya que influye en la productividad e ingresos de la población. Además estudios recientes resaltan el efecto en el bienestar de las familias, las empresas y el rendimiento de una economía (Alsan et al, 2007). Evidencia que ha afectado el enfoque de política en los últimos años en varios países. Como resultado se puede apreciar que de 8 Objetivos del Milenio planteados por las Naciones Unidas en el año 2000, 3 se encuentran directamente relacionados con la salud. Estudios sugieren una estrecha relación entre crecimiento económico y una población más saludable. Además, y aún más importante, se ha encontrado que la mejoría de la salud tiene un impacto positivo y considerable en el crecimiento económico, la riqueza de la población y la igualdad económica. (Bloom & Canning , 2009)

Conociendo el efecto de la salud en la economía, es de interés estudiar a mayor profundidad las características que llevan a la población a presentar problemas de salud y posteriormente demandar servicios de salud. Mediante modelos econométricos se puede conocer una relación causal entre las variables analizadas y la demanda de salud, además se puede cuantificar el efecto de cada variable, permitiéndose categorizarlas por el tamaño de dicho efecto, lo cual es de utilidad para el tomador de decisiones de política pública al momento de asignar recursos. Con dichos insumos en ideal se podría construir una política que maximice los beneficios de la salud en la economía del país. Para alcanzar dicho fin se utiliza la Encuesta Nacional de Salud y Nutrición 2011 – 2013 (ENSANUT 2011-2013). Donde se encuentra que el 42.83% de los ecuatorianos tuvieron algún problema de salud el mes anterior a la entrevista, lo que representaría un total de 8.8 millones de personas. Del cual el 21.66% del total poblacional asistió a un centro de salud, consultorio privado, etc.; y, un 14.16% se automedicó (Tabla No. 1). Igualmente se puede apreciar en la Tabla No. 1 que 6.16% de la población, pese a presentar un problema de salud decide no realizar ninguna acción. Cabe recalcar que a partir de la ENSANUT no se puede obtener información sobre las razones para la automedicación.

Tabla No. 1
Población con problemas de salud y acción realizada frente al problema de salud

Categoría	Tuvo algún problema de salud en los últimos 30 días?				
	Total	Proporción	Error Estandar	Intervalo de confianza 95%	
Si	8.862.482	42,83	0,39	42,07	43,59
<i>Acción realizada</i>					
Visitó un centro de salud, consultorio, etc.	4.482.134	21,66	0,53	20,84	22,50
Obtuvo atención en casa	95.382	0,46	0,12	0,37	0,58
Se automedicó	2.991.834	14,46	0,50	13,79	15,15
Tuvo que internarse	19.427	0,09	0,04	0,07	0,13
No hizo nada	1.273.705	6,16	0,35	5,76	6,57
No	6.638.467	57,17	0,39	56,41	57,93

Fuente: ENSANUT 2011-2013

Elaboración: Diego Villacreses

Con el fin de analizar las razones para no tomar acciones frente al problema de salud se puede observar la Tabla No. 2, apreciándose que del total de personas que reportan encontrarse en esta categoría un 37.77% considera que la enfermedad o accidente es leve, un 20.1% que conoce la enfermedad, mientras que un 15.79% reporta no poder acceder a servicios de salud por razones económicas.

Tabla No. 2
Razón para no tomar acciones frente al problema de salud

Categoría	Razón por la cual no tomo acción frente a la enfermedad				Intervalo de confianza 95%
	Total	Proporción	Error Estandar		
Enfermedad o accidente leve	481.040	37,77	1,16	35,53	40,06
No tuvo tiempo	123.060	9,66	0,62	8,50	10,96
Centro de atención lejano	11.598	0,91	0,17	0,63	1,32
Muy costoso	14.085	1,11	0,20	0,78	1,57
Atención de mala calidad	29.512	2,32	0,27	1,84	2,91
Falta dinero	201.098	15,79	0,95	14,01	17,75
No pudo dejar la casa o los niños	10.828	0,85	0,25	0,48	1,51
Conoce la enfermedad	256.013	20,10	0,93	18,33	21,99
No puede dejar de trabajar	35.562	2,79	0,32	2,23	3,49
Otro	110.910	8,71	0,66	7,50	10,09

Fuente: ENSANUT 2011-2013

Elaboración: Diego Villacreses

Tras el análisis realizado a la información presentada por la ENSANUT se puede apreciar un problema de salud nacional no atendido, el cual se encuentra presente tanto en las personas que decidieron auto medicarse como no atenderse por razones diferentes a la levedad de la enfermedad.

Es de importancia utilizar la ENSANUT 2011-2013 ya que es la única fuente de información que permite relacionar la demanda de servicios de salud con características comportamentales y socio-demográficas de las personas a nivel de microdata, permitiendo realizar modelos econométricos que permitan caracterizar esta relación. Por otro lado y como fuente para contrastar la información se tienen los registros administrativos del INEC. Donde se encuentran las Actividades y Recursos de Salud que presentan a nivel agregado por institución de salud información del total de camas, profesionales, recursos de laboratorios, atenciones ambulatorias y hospitalizaciones (Instituto Nacional de Estadísticas y Censos, 2013a). Además se dispone de la información de Egresos Hospitalarios donde se registra únicamente nombre, lugar de residencia y causa de hospitalización (Instituto Nacional de Estadísticas y Censos, 2013b), esta información representa únicamente la demanda efectiva de servicios de salud en relación a internamiento. Lo que dificulta la planificación al desconocer la demanda potencial de servicios de salud en ámbitos como consultas externas tanto públicas como privadas.

Por las razones anteriormente expuestas se decide proponer la presente disertación. Pues a partir de un modelo de determinantes se puede conocer las características que influyen en la demanda de servicios de salud; y además la intensidad con la que lo hacen. A partir de un

modelo correctamente especificado se pueden considerar los estimadores de bondad de ajuste del mismo para conocer su capacidad predictiva, y proponer un marco de clasificación de personas más riesgosas de sufrir problemas de salud no leves y por consiguiente representar demanda de servicios.

Según Samuelson (1954) se pueden asumir dos categorías de bienes: bienes privados ordinales y bienes de consumo colectivo. Este último tiene dos características: no excluible (no se puede evitar que una persona consuma si esta no paga dicho bien) y no rival (el consumo de una persona no evita que otras la consuman simultáneamente). Samuelson no asume la existencia de un mecanismo que asigne los bienes colectivos óptimamente. De hecho para demostrar la existencia de un óptimo social, asume que cada individuo posee un conjunto de preferencias bien definidas sobre las cuales decide su consumo de bienes tanto privados como públicos. Para relajar dicho supuesto se asume la existencia de bienes cuasi-públicos donde el estado decidirá la oferta de dichos bienes en función de la demanda agregada (McConnell et al, 2012). Como se encuentra en la literatura, en general se asume el conocimiento o capacidad de estimar la demanda para establecer la oferta y así encontrar el punto de equilibrio que llegue a un óptimo social.

Como se expuso, la salud de las personas es fundamental para el crecimiento económico y el bienestar individual. Partiendo de la necesidad de un sistema externo al mercado que asigne de forma óptima recursos de salud, se plantea estimar la demanda y sus determinantes. Por esta razón la presente disertación propone el uso de herramientas estadísticas y econométricas para establecer los determinantes de la demanda de servicios de salud a nivel de microdata.

La presente disertación se estructura en tres secciones, la primera presenta una recopilación de la literatura necesaria para comprender la demanda de salud, donde se incluye el modelo teórico de demanda de Salud de Grossman y la aplicación empírica realizada por el mismo, se analizan las herramientas econométricas mediante las cuales puede caracterizar la demanda de salud, como son la Regresión Lineal, los Modelos de Variables Binarias y los Modelos de Conteo. Además se analizan las consideraciones teóricas que se deben tener al momento de estudiar los coeficientes de cualquier regresión, tomando en cuenta una posible relación causal y las condiciones para afirmar dicha relación. Finalmente, tomando en cuenta la subsección anterior, se revisan los determinantes teóricos y empíricos de la demanda de salud.

Metodología de Investigación

La presente disertación tiene por objetivo estudiar la demanda de salud, por lo que se utiliza la única fuente de información pública mediante la cual se pueden explorar dicho objetivo, la ENSANUT 2011-2013. Donde se encuentra información sobre la presencia enfermedades en la población, con desagregación por tipo de enfermedad y cantidad de tiempo que paso enfermo. Por ser información a nivel de microdata (se tienen los datos a nivel de persona) se puede plantear un modelo de regresión en base a las características individuales de los encuestados y se puede inferir los resultados a nivel nacional gracias a los factores de expansión.

Ya que las variables a considerarse para el modelo deben provenir de una explicación teórica del fenómeno con el fin de conocer con mayor precisión la relación entre dichas variables y el fenómeno estudiado se presentan los estudios realizados sobre el tema. Posteriormente se analizan los modelos econométricos a usarse y finalmente se presentan los resultados de las regresiones con sus respectivos análisis e interpretaciones.

Metodología que genera las siguientes preguntas y objetivos de la presente investigación:

Pregunta general

¿Cuáles son las características de la demanda de servicios de salud en el Ecuador?

Preguntas específicas

¿Cuáles son las condiciones específicas de la demanda de salud en el Ecuador según la ENSANUT 2011-2013?

¿Cuáles son los determinantes de la demanda de salud en el Ecuador según la ENSANUT 2011-2013?

¿Se puede considerar a dicho modelo de determinantes un modelo de predicción?

Objetivo General

Caracterizar la demanda de salud encontrando las variables observables que determinan a la misma.

Objetivo Específico

Analizar y estimar las condiciones de la demanda de salud en el Ecuador en base a la ENSANUT 2011-2013 y comparar dicha información con los Registros Administrativos.

Encontrar y analizar las variables observables que determinan la demanda de servicios de salud para el caso ecuatoriano, buscando generar un modelo econométrico correctamente especificado, donde se encontraría la relación real entre dichas variables y la demanda de servicios de salud.

Mediante el uso de indicadores de bondad ajuste determinar si los modelos de determinantes propuestos pueden considerarse como modelos de predicción.

Fundamentación Teórica

Bienes Públicos

Samuelson (1954) es el primero en proponer un óptimo social en función de dos tipos de bienes: bienes colectivos o públicos y bienes privados. Donde se define a los bienes privados como un vector (X_1, \dots, X_n) repartidos entre los individuos $(1, 2, \dots, i, \dots, s)$ y los bienes públicos como $(X_{n+1}, \dots, X_{n+m})$, los cuales pueden ser utilizados o reportar utilidad a todos los individuos, sin que este consumo disminuya la utilidad de cualquier otro individuo que vaya a consumir dicho bien. Además se asume que ningún individuo i podrá ser excluido del consumo del bien X_{n+j} pese a no pagar por este. Se asume para cada individuo i una utilidad convexa u^i en función de los bienes tanto privados como públicos que consume. No se asume un estado que asigne los bienes públicos, por lo que se propone que cada individuo i escoja los bienes públicos y privados que maximicen su utilidad individual. De esta forma la sumatoria de las utilidades individuales expresará el óptimo social. Finalmente Samuelson recomienda el uso de las ramas política del bienestar ampliamente estudiada por Arrow, Duncan y otros.

La definición de bienes públicos generada por Samuelson es sumamente restrictiva, de hecho se encuentra que existen pocos casos reales que cumplen con dichas propiedades. Los servicios de salud pública no son este tipo de bienes colectivos. De hecho existe una amplia discusión sobre qué tipo de bien debe ser considerado el acceso a salud (Culyer y Newhouse, 2000), existiendo numerosos modelos y evidencia empírica soportando tanto una concepción de la salud como privada, pública o mixta. Si bien en años recientes se ha presenciado una inclinación de los estados por la salud pública. El propósito de la presente disertación no se centra en proponer el mejor de modelo de acceso a salud.

Se puede evidenciar que en general se asume información perfecta en los mercados (Stiglitz, 2000). Por lo que tanto consumidores como productores poseerán el conjunto completo de información del mercado. Sabiendo que este no es el caso para compradores o consumidores, en la presente disertación se busca caracterizar y predecir la demanda de servicios de salud.

Modelos de Predicción

Modelo de predicción es un término general que se refiere a un conjunto de técnicas, matemáticas, estadísticas y econométricas; cuyo fin específico es conocer de la forma más precisa posible y con anterioridad el comportamiento de un fenómeno (Elkan, 2013). La ciencia económica utiliza tanto modelos estadísticos, matemáticos como econométricos para predecir. Usualmente estas técnicas parten de un marco teórico desarrollado dentro de la economía. (Wooldridge, 2010).

Con el fin de predecir un fenómeno económico se pueden utilizar dos enfoques. El primero se basa en el uso de algún modelo de predicción de forma atórica, donde las relaciones entre el fenómeno estudiado y la información usada para su predicción no tiene justificación dentro de la ciencia económica. Un ejemplo del mismo puede ser un modelo logístico para la predicción del riesgo financiero de los clientes de un banco, construido a partir de las variables más

significativas estadísticamente. Una crítica a este tipo de modelo se da cuando la información se actualiza, la matriz de varianza y covarianza de los estimadores del modelo cambiará, modificando así la significancia de las variables, lo cual a su vez cambiará las variables a tomarse en cuenta por el modelo independientemente de cualquier teoría económica o financiera.

El segundo enfoque se basa en la predicción a partir de la aplicación de un modelo teórico. Este enfoque se usa cuando el fenómeno de interés ha sido estudiado previamente en la teoría económica. El primer paso de un modelo de predicción bajo este marco será la caracterización correcta del fenómeno social de interés, argumentando teóricamente la causalidad de un conjunto de variables sobre el fenómeno; y, su posterior demostración empírica mediante el uso de una técnica econométrica o estadística. En el caso de no existir causalidad estadísticamente demostrable se podría argumentar la no aplicabilidad del modelo teórico para el caso específico estudiado. Una vez construido un modelo teóricamente correcto se procede a la predicción del fenómeno de dentro de un conjunto previamente definido de información para comprobar la capacidad predictiva del modelo mediante estadísticos de ajuste. En general no existe una demostración definitiva de la superioridad de un modelo frente al otro. (Elkan, 2013)

La presente Disertación se basará en el segundo enfoque. Específicamente se utilizarán técnicas econométricas, ya que resultan ser las más adecuadas para el fenómeno estudiado (Wooldrige, 2010). Se presentaran varias técnicas econométricas basadas en modelos lineales y no lineales. Como primera instancia se caracteriza al fenómeno a estudiar como una variable Y , la cual presentará i observaciones, por lo que quedará definida como Y_i . Esta variable quedará explicada por un conjunto $X_i(x_{1i}, x_{2i}, x_{3i}, \dots, x_{pi})$ de variables independientes, mediante la esperanza condicional de X_i sobre Y_i :

$$E(Y_i|X_i) = X_i'\beta + u_i$$

Método que se denomina Regresión Lineal Simple. Donde con el fin de poder estimar un efecto causal del conjunto X_i sobre Y_i se asume exogeneidad de las variables X_i en relación al término de error e_i : $E(X_i|e_i) = 0$. El supuesto de exogeneidad significa que ninguna variable x_i del modelo presenta correlación con el término de error e_i . Dado que no existe un test para comprobar el cumplimiento de este supuesto, (sin asumir ya su rompimiento o endogeneidad y una corrección del mismo) la construcción del modelo y la argumentación de la exogeneidad se dan de manera teórica (Angrist & Pischke, 2009). La generación de un modelo bajo dichos parámetros permitirá calcular el efecto que tiene cada variable x_j sobre el fenómeno estudiado, donde el parámetro β_j representará el efecto sobre la variable Y de un cambio marginal de la variable x_j . Cabe recalcar la diferencia entre regresión poblacional y regresión muestral, donde la regresión poblacional obtendrá los parámetros u efecto poblacionales β_j . Ya que generalmente se trabaja con muestras, se utiliza una estimación de $\beta_j \approx \hat{\beta}_j$. (Wooldrige, 2010)

Un ejemplo una regresión correctamente especificada puede darse en base al ingreso en dólares. Donde una de las variables dependientes sea el género de la persona, con $x=1$ si la persona es mujer; y, con un estimador $\hat{\beta}$ estadísticamente significativo. Expresará que las

mujeres ganan $\hat{\beta}$ más que los hombres pese a mantener todas las demás características iguales.

Una vez construido un modelo correctamente especificado, se puede proceder a la predicción de la variable Y_i mediante la estimación de la misma:

$$\hat{Y}_i = X_i' \hat{\beta}$$

La presente disertación tratará a la demanda de salud en dos ámbitos, como variable dicotómica (si la persona se enfermó o no) y como variable de conteo (cantidad de días hospitalizado) Para el caso dicotómico:

$$Y_i \begin{cases} 1 & \text{si la persona se enfermó} \\ 0 & \text{caso contrario} \end{cases}$$

Para el caso de variable de conteo:

$$Y_i \begin{cases} n & \text{días hospitalizado} \\ 0 & \text{caso contrario} \end{cases}$$

Se pueden usar varias especificaciones con el fin de modelar los determinantes al tomar en cuenta este tipo de variables. Para variables dicotómicas se cuenta con los modelos Logístico, Probabilístico y de Probabilidad Lineal. En el caso de las variables de conteo se puede modelar: regresión de Poisson, Binomial Negativa, entre otros. A continuación se presentan los diversos modelos que se pueden utilizar para la caracterización y predicción de variables binarias y ordinales.

Modelos de Variables Binarias

Un Modelo de Variable Binaria se utiliza cuando la variable dependiente Y adquiere únicamente valores 0 o 1. Este tipo de variables se observan en fenómenos en como el uso de servicios de salud, compra de seguros de salud, fumar, etc. (Jones, 2000). En el contexto de una regresión cuando un modelo tiene como variable dependiente una variable dicotómica, los errores también seguirán dichos valores. Además la varianza de los estimadores $\hat{\beta}$ serán no constantes (heterocedasticidad).

En general un modelo de regresión queda especificado como la esperanza condicional de Y en relación a un vector de variables X . Para el caso en el que se tiene una variable binaria como dependiente la esperanza condicional queda expresada como:

$$E(Y_i|X_i) = P(Y_i = 1|X_i) = F(X_i)$$

Donde $P(Y_i = 1|X_i)$ se referirá a la probabilidad de que la variable Y_i sea igual a 1 dado el vector X_i . En el contexto de una variable dependiente binaria la esperanza condicional es equivalente a

una probabilidad condicional. La cual se expresará como una función lineal o no lineal de X_i ; $F(X_i)$.

Modelo de Probabilidad Lineal

El modelo de probabilidad lineal es el más sencillo de los modelos de respuesta binaria y fue el primero en desarrollarse. Su especificación se puede derivar de la resolución de la ecuación anterior, dada la característica lineal del modelo:

$$P(Y_i = 1|X_i) = F(X_i) = X' \beta$$

El modelo de probabilidad lineal presenta un error de consistencia lógica al predecir valores fuera del rango $[0,1]$, lo cual viola el primer axioma de la probabilidad (Wasserman, 2010):

La probabilidad es positiva y menor o igual a 1:

$$0 \leq p(A) \leq 1$$

Además el término de error seguirá una distribución bimodal alrededor de los valores de una variable dicotómica ósea: 0 y 1. Además:

$$Var(Y|X) = X\beta(1 - X\beta)$$

Lo cual revela la presencia de heterocedasticidad de los errores a menos que cada elemento del vector de coeficientes sea cero $\beta(\beta_1, \beta_2, \dots, \beta_k)$. Por lo que al utilizar un modelo de probabilidad lineal se recomienda utilizar errores estándares robustos con el fin de realizar correctamente los correspondientes test de hipótesis (Wooldrige, 2002).

Modelo Logístico o Logit

El Logit es un modelo de regresión para variables binarias. Puede comprenderse a partir de una interpretación de variable latente:

$$Y_i = \begin{cases} 1 & \text{si } Y_i^* > 0, \\ 0 & \text{si } Y_i^* \leq 0 \end{cases}$$

Donde la variable dependiente adquirirá un valor igual a 1 si la variable latente y_i^* alcanza un valor mayor a 0. La variable latente quedará especificada de la siguiente forma:

$$Y_i^* = X_i \beta + u_i$$

Para derivar un modelo de regresión en base a los términos anteriores, se propone la probabilidad condicional de $Y_i = 1$ dado X_i : $P(Y_i = 1|X_i)$. Sabiendo que Y_i será igual a 1 cuando la variable latente Y_i^* sea mayor a 0, la probabilidad condicional quedará expresada como:

$$P(Y_i = 1|X_i) = P(Y_i^* > 0|X_i) = P(X_i\beta + u_i > 0)$$

Mediante propiedades de las funciones de distribución acumulada, donde se enuncia que la probabilidad de que una variable aleatoria x sea mayor o igual a un valor X , se formula: $F(x) = P(X \leq x)$. Por consiguiente asumiendo que el término de error e_i siga una distribución logística el modelo queda especificado de la siguiente manera:

$$P(Y_i = 1|X_i) = F(X'\beta) = \Lambda(X'\beta) = \frac{e^{X'\beta}}{1 + e^{X'\beta}}$$

A diferencia de una regresión lineal los coeficientes de un modelo logit (y el probit) no tienen la interpretación de un efecto de la variabilidad de la variable x estudiada sobre la variable independiente (Cameron & Trivedi, 2009). Dado que los coeficientes β presentan un comportamiento no lineal frente a la variabilidad de las x , dicha conducta se ve explicada por la siguiente ecuación:

$$\frac{\delta P(Y = 1|X)}{\delta X} = \beta F'(X'\beta) = \beta \lambda(X'\beta)$$

En el modelo de regresión lineal la derivada de Y frente al vector de variables X es igual al coeficiente β . En el caso de un modelo logístico o probabilístico dicha derivada es igual a una ecuación no lineal (además $F'(X'\beta)$ presenta un comportamiento no lineal frente a variaciones en X), la cual se denomina efectos marginales. El efecto marginal se debe evaluar en punto de la distribución de las X , este es usualmente el promedio o la mediana. De los efectos marginales se puede obtener una interpretación de efecto promedio de la variabilidad de X frente a Y (Jones, 2005).

Modelo Probit

Bajo los mismos parámetros de un modelo de regresión para variables binarias expuesto anteriormente. Al asumir una distribución probabilística o normal para el término de error, el modelo quedará especificado de la siguiente forma:

$$P(Y_i = 1|X_i) = F(X'\beta) = \Phi(X'\beta)$$

Donde Φ se refiere a la distribución acumulada de probabilidad normal o Gaussiana (Greene, 2010). Ya que los coeficientes β no representan el efecto de un cambio marginal en X , se tiene el mismo problema que en el modelo logístico, por lo que se utilizan los efectos marginales evaluados en el promedio o mediana:

$$\frac{\delta P(Y = 1|X)}{\delta X} = \beta F'(X'\beta) = \beta \varphi(X'\beta)$$

A continuación en la Tabla No. 3 se enuncian las ventajas y desventajas generales de los modelos logístico, probabilístico y lineal de probabilidad. Como puede apreciarse cada modelo presenta ventajas bajo las cuales es razonable utilizar a los mismos. Para los fines de la

presente disertación se decide utilizar los modelos logísticos, donde se evita la inconsistencia lógica de estimar valores para la variable independiente fuera del intervalo comprendido entre 0 y 1. Además no se utilizarán variables instrumentales por lo que no es un problema la limitación del uso de variables instrumentales continuas.

Tabla No. 3		
Ventajas y desventajas de los modelos probit, logit y lineal de probabilidad		
Lineal de Probabilidad	Logit	Probit
<i>Ventajas</i>		
Los coeficientes son de fácil interpretación.	Se puede estimar el Ratio de Odds	Los valores de predicción se encuentran acotados entre [0,1]
La estimación es sencilla.	Los valores de predicción se encuentran acotados entre [0,1]	
Se puede corregir un problema de endogeneidad mediante variables instrumentales no continuas.		
<i>Desventajas</i>		
Los valores predichos de Y sobrepasan el intervalo [0,1]	Se puede corregir un problema de endogeneidad únicamente mediante variables instrumentales continuas.	Se puede corregir un problema de endogeneidad únicamente mediante variables instrumentales continuas.
Los errores no siguen una distribución normal	Se asume una probabilidad igual entre las categorías de la variable dependiente ($\pi=0.5$)	No se puede calcular el Ratio de Odds
		La estimación es computacionalmente difícil. Se asume una probabilidad igual entre las categorías de la variable dependiente ($\pi=0.5$)

Fuente: Cameron & Trivedi (2009), Wooldridge (2002)

Elaboración: Diego Villacreses

Modelos de conteo

En la presente disertación se modelará la cantidad de días de hospitalización con el fin de caracterizar la relación entre dicha variable y variables socio-demográficas, económicas y de comportamiento. Para dicho fin se utilizarán modelos de regresión de conteo, ya que la cantidad de días es una variable discreta $N=\{0,1,2,\dots,n\}$ no negativa, que se diferencia de las variables

categorías y ordinales ya que esta mide el número de ocurrencias de un evento, definición de una variable de conteo (Karazsia & van Dulmen, 2008). Dentro de los modelos de conteo se encuentran la regresión de Poisson, la Binomial Negativa y los Modelos Zero-Inflados, entre otros. A continuación se detallarán los principales modelos.

Regresión de Poisson

La regresión de Poisson asume que la variable dependiente se generó mediante un proceso generador de datos de una distribución Poisson, la que se pudo generar:

- Conteo derivado de un Proceso Puntual: se refiere a conteos observados directamente, como por ejemplo el número de faltas en el trabajo o el número de accidentes aéreos.
- Conteo derivado de Data Continua: se obtiene por la categorización de una variable continua latente, como por ejemplo el ranking de crédito. Hay que tomar en cuenta que este tipo de variables pueden modelarse también con un modelo probit o logit ordenado.

Siendo una distribución de Poisson una distribución de probabilidad discreta que corresponde a un subconjunto aleatorio y contable que cumple con el teorema de aditividad (Kingman, 1993). Una variable Y presentará una distribución Poisson con una tasa de intensidad $\mu > 0$ que representa el número de veces que se espera suceda el fenómeno, y una exposición t definida como la cantidad de tiempo en la que suceden los eventos, generalmente se asume a la misma como 1 (Cameron & Trivedi, 1998), de tal forma que la distribución queda definida:

$$Pr[Y = y] = \frac{e^{-\mu t} (\mu t)^y}{y!}, y = 0, 1, 2, \dots$$

Siendo el primer momento de la distribución:

$$E[Y] = \mu$$

Una propiedad de la distribución de Poisson es la equidispersión, la que se refiere a la igualdad entre la varianza y la media, propiedad que pocas veces se observa en la información de la vida real. Cuando la varianza es más grande que la media se denomina sobre-dispersión.

El modelo de regresión de Poisson se deriva de la distribución con el mismo nombre, donde el parámetro de intensidad μ dependerá de un conjunto de covariables X . Si el modelo se encuentra correctamente especificado únicamente en base al vector X se denomina regresión de Poisson estándar. Por otra parte, si se considera que la regresión incluye variables aleatorias no observables se tiene un modelo de regresión Poisson Mixto. La regresión nace de la esperanza condicional $E[Y_i|X_i]$; y, ya que la FEC es no lineal:

$$E[Y_i = y_i|X_i] = g(Y_i = y_i, X_i) = \frac{e^{-\mu_i} \mu_i^{y_i}}{y_i!}$$

Donde se puede proponer una tasa de intensidad para el modelo Log-linealizado:

$$\mu_i = \exp(X_i\beta)$$

Cabe recalcar una propiedad importante de la regresión de Poisson, ya que $E[Y_i|X_i] = V[Y_i|X_i]$ el modelo será intrínsecamente heterocedástico. (Cameron & Trivedi, 1998).

Test de sobredispersión

Cuando una distribución no cumple con el supuesto de equidispersión la regresión de Poisson presentará un error parecido a la violación del supuesto de homocedasticidad en una regresión lineal (Cameron & Trivedi, 1998). La forma más sencilla de aproximar una estimación de sobredispersión es calculando el promedio y la varianza de la variable dependiente; y, comparándolas, si bien en este caso la varianza incondicional de Y debería ser al menos el doble de la media incondicional, ya que al incluir regresores en el modelo la varianza condicional caerá considerablemente (casi a la mitad de lograrse un R² igual a 0.5).

La prueba más usada para estimar la sobredispersión de una regresión de Poisson es mediante el uso de una regresión Binomial Negativa, ya que esta es un caso general de la regresión de Poisson donde $\alpha > 0$. Por lo que se buscará aceptar la hipótesis nula:

$$H_0: \alpha = 0$$

El problema de la estimación causal

Entre los objetivos de una regresión se encuentra la caracterización de un fenómeno social o económico. Donde al cumplir con el supuesto de exogeneidad del vector de variables X en relación al término de error se puede conocer un efecto causal de cada variable del vector X sobre la variable Y.

Bajo dicha óptica la interpretación de los coeficientes β en una regresión lineal o los efectos marginales en un modelo de respuesta binaria no lineal, sería el de la variación presentada por Y frente un cambio marginal en una de las variables del vector X manteniéndose el resto del vector invariante (Wooldrige, 2010).

Conocer el efecto causal de una variable sobre el fenómeno estudiado es el problema más importante en ciencias sociales. Al comprender la causalidad de una variable A sobre la variable B, se puede realizar predicción en escenarios donde A cambie de alguna forma. Pudiendo así planificar el efecto deseado en B mediante la manipulación de A. Para comprender la causalidad de un conjunto de variables $X(x_1, x_2, \dots, x_p)$ sobre Y se debe garantizar la exogeneidad del modelo (Angrist & Pischke, 2009) (Greene, 2008) (Wooldrige, 2010). La capacidad de comprender el efecto de X sobre Y permitirá no solo predecir; permitirá además tomar decisiones sobre las variables x_j que se puedan manipular. Como por ejemplo tras comprender correctamente el efecto que tiene la cantidad de alumnos en un aula escolar sobre el rendimiento de los mismos, se puede determinar la cantidad optima de alumnos para potenciar su aprendizaje (Angrist & Pischke, 2009).

Para garantizar la exogeneidad en una regresión existen dos enfoques. El primero se da cuando existe la posibilidad de asignar aleatoriamente la variable cuyo efecto se desea conocer. El segundo se da con data observacional, ya que ninguna variable fue asignada de forma aleatoria, se deberá construir un modelo con un conjunto de variables X de tal forma que se cumpla el supuesto de exogeneidad. La forma de escoger dicho conjunto de variables viene de la teoría económica, ya que no existe un test estadístico que pueda determinar si un modelo es endógeno sin antes argumentar la endogeneidad del mismo mediante una variable instrumental (Greene, 2008). La segunda forma de garantizar exogeneidad será usada en la presente Disertación ya que no existe data experimental sobre el tema en el Ecuador. Un modelo con dichas características se denominará un modelo correctamente especificado.

Ya que un modelo correctamente especificado puede ser usado para la toma de decisiones, por ejemplo en política pública o el curso de acción de una empresa. Se debe comprender más profundamente la exogeneidad ya que su cumplimiento como supuesto en términos prácticos es complejo.

Exogeneidad

La exogeneidad quedará expresada en términos generales de la siguiente forma:

$$E[\varepsilon_i | X_i] = 0$$

Lo que significa que la esperanza matemática del término de error en la observación i debe ser independiente del vector de variables X para toda observación i , incluyéndose a sí misma (Greene, 2008). Y dado que: $Y_i = E[Y_i | X_i] + \varepsilon_i$. Se puede demostrar la fórmula anterior de la siguiente forma:

$$E[\varepsilon_i | X_i] = E\{Y_i - E[Y_i | X_i] | X_i\} = E[Y_i | X_i] - E[E[Y_i | X_i] | X_i] = E[Y_i | X_i] - E[Y_i | X_i] = 0$$

La anterior ecuación se demuestra gracias a la Ley de Esperanzas Iteradas: $E\{E[Y_i | X_i]\} = 0$. Y por la propiedad de descomposición de las mismas.

Se decide usar la esperanza condicional para expresar la relación entre X e Y , ya que esta es el mejor predictor bajo la óptica del problema de predicción del Mínimo Error Cuadrado Promedio (Angrist & Pischke, 2009):

$$E[Y_i | X_i] = \arg \min_{m(X_i)} E[(Y_i - m(X_i))^2]$$

Y la estrecha relación entre la esperanza condicional y la regresión y la regresión lineal puede ser comprendida en relación a tres Teoremas (Angrist & Pischke, 2009). El primero se denomina el Teorema de la Función de Esperanza Condicional (FEC) Lineal:

A partir de la minimización de los errores al cuadrado ($\beta = \arg_b \min E[(Y_i - X_i' b)^2]$). De donde se obtienen las Condiciones de Primer Orden:

$$E[X_i(Y_i - X_i' b)] = E[X_i(Y_i - E[Y_i | X_i])]$$

Si se asume que $E[Y_i|X_i] = X_i'\beta$. Se resuelve: $\beta = E[X_i'X_i]^{-1}E[X_i'Y_i]$.

El siguiente es el Teorema de la Mejor Predicción Lineal, donde se enuncia que al igual que la FEC es el mejor predictor de Y_i dado X_i para toda función de X_i , la regresión lineal es la mejor aproximación posible para funciones lineales. El tercero se denomina Teorema de la Regresión FEC, el cual enuncia que pese a un FEC no lineal la regresión es la mejor aproximación lineal posible.

Bajo el tercer teorema se puede argumentar fuertemente el uso de una regresión para un problema no lineal como por ejemplo un modelo logístico. Si bien no se obtiene la información exacta sobre la distribución analizada a nivel de Y_i , cuando el interés es la distribución de Y el tercer teorema enuncia que una regresión es suficientemente buena. Para el caso de la presente disertación el objetivo es predecir en áreas pequeñas la demanda de servicios de salud, para lo cual inicialmente se pretende comprender mediante una FEC la demanda de servicios de salud. Para dichos fines, según el tercer teorema una regresión lineal simple es suficiente (Angrist & Pischke, 2009).

Las esperanzas condicionales fueron analizadas en la presente subsección dada su importancia para comprender correctamente el concepto de exogeneidad y en el proceso se justifica robustamente el uso de una regresión lineal mediante Mínimos Cuadrados Ordinarios en la estimación de Áreas Pequeñas de la presente disertación.

El cumplimiento de la exogeneidad garantiza la estimación de los verdaderos valores de β en una regresión lineal poblacional. Y frente a una regresión muestral garantizará las propiedades de insesgamiento y consistencia (Wooldridge, 2002). La exogeneidad se incumple en tres situaciones:

- Variable omitida
- Error de medida
- Simultaneidad

Variable Omitida

Asumiendo que el fenómeno Y pueda ser explicado por una FEC lineal de la siguiente forma $E(Y_i|X_i, q_i)$. El modelo únicamente podrá ser estimado si $cov(q, X) = 0$. En el caso de existir alguna correlación entre una variable del vector de X_i y q_i el modelo será endógeno. Ya que q_i no puede ser estimado, queda incluido en el error $u_i = q_i + v_i$. Al existir una correlación entre q_i y X :

$$q_i = X_i\delta$$

Y dado el modelo:

$$Y_i = X_i'\beta + u_i = X_i'\beta + q_i + v_i$$

Cuando existe alguna correlación entre q_i y el conjunto de variables independientes, se puede demostrar el sesgamiento de los coeficientes estimados mediante $E[Y_i|X_i]$:

$$Y_i = X_i'(\beta + \delta) + v_i$$

De tal forma que se demuestra que frente a la correlación de la variable omitida con X_i los coeficientes estimados no son los verdaderos β , sino $(\beta + \delta)$. Cabe recalcar que hasta el momento se ha analizado únicamente el modelo de regresión poblacional. Al trabajar con una muestra los coeficientes β se estiman mediante $\hat{\beta}$.

El vector $\hat{\beta}$, debe cumplir con dos propiedades: insesgadez y consistencia. La insesgadez se refiere a:

$$E(\hat{\beta}) = \beta$$

Dicha propiedad enuncia que la esperanza matemática del estimador es igual al parámetro. Lo cual puede ser comprendido como el promedio de todas las estimaciones $\hat{\beta}$ provenientes de todos los muestreos posibles de una población, en promedio serán igual a β . La siguiente propiedad es la consistencia:

$$\text{plim } \hat{\beta}_j = \lim_{n \rightarrow \infty} \sqrt{n}(\beta_j - \hat{\beta}_j) = 0$$

La cual enuncia que el estimador $\hat{\beta}$ se aproximará a β , a medida que el tamaño de la muestra se acerca al infinito.

En una regresión que presenta variable omitida correlacionada a X_i la convergencia para todo j será:

$$\text{plim } \hat{\beta}_j = \beta_j + \delta_j$$

Y la esperanza $E(\hat{\beta} + \delta) = \beta + \delta$. Por lo tanto se puede concluir que una regresión muestral con variable omitida correlacionada con X_i es inconsistente y sesgada. Tanto para el caso de una regresión muestral como poblacional los coeficientes reportados no reflejarán el verdadero efecto de X_i sobre Y_i . Una regresión endógena mediante variable omitida puede ser solucionada mediante el uso de una variable *proxy*.

Error de Medida

El error de medida tiene una estructura estadística similar a la variable omitida, si bien conceptualmente son profundamente diferentes (Wooldrige, 2002). El error de medida se da cuando la información observada de Y_i o x_{ji} mediante las variables Y_i^* o x_{ji}^* no es la información real:

$$Y_i^* \neq Y_i$$

$$x_{ji}^* \neq x_{ji}$$

Dicho problema se da usualmente cuando los agentes económicos tienen incentivos a ocultar o modificar información reportada. Como por ejemplo en reporte de ingresos o ahorro los agentes tienden a modificar sus respuestas. El error de medida puede clasificarse según al tipo de variable que afecte: variable dependiente o variable independiente.

Error de medida de variable dependiente

Se da cuando la variable Y_i^* observada no corresponde al valor real de Y_i . Puede expresarse de la siguiente forma:

$$Y_i^* - Y_i = \zeta_i$$

Este tipo de error de medida solamente representará un problema cuando ζ_i se encuentre correlacionado con alguna variable explicatoria del vector de variables X_i que garantice un modelo correctamente especificado. El modelo de regresión se puede plantear:

$$Y_i^* = X_i\beta^* + u_i$$

En el caso de: $E[X_i|\zeta_i] = 0$, se podrá garantizar: $\beta^* = \beta$; donde β se obtiene de la regresión MCO con variable dependiente Y_i :

$$Y_i = X_i\beta + u_i$$

Lo cual garantiza estimadores insesgados y consistentes en la regresión muestral; y, coeficientes poblacionales iguales entre la regresión con la variable Y_i (sin error de medida) frente a la regresión de Y_i^* (con error de medida). Significando que los coeficientes β^* representaran el verdadero efecto de X_i frente al fenómeno estudiado independientemente del error de medición. Se puede encontrar en el Anexo No. 1 un código en STATA donde se demuestra mediante un proceso generador de datos basado en simulaciones Monte Carlo dicha afirmación. La única problemática presentada por una regresión con variable dependiente con error de medición es que los errores estándares de los coeficientes β^* son más grandes que sus respectivos errores estándares de β .

Error de medida de variable independiente

Dada la regresión:

$$Y_i = X_i\beta + \varepsilon_i = \beta_0 + \beta_1x_{1,i} + \beta_2x_{2,i} + \dots + \beta_kx_{k,i} + u_i$$

Donde se pueden observar todas las variables $x_{j,i}$ sin error de medida, esta regresión poblacional presentaría coeficientes insesgados, que reflejarían el verdadero efecto de cada variable $x_{j,i}$ sobre Y_i . Cuando en la regresión se incluye una variable $x_{k,i}^* = x_{k,i} + \vartheta_i$ con un error de medida se estimaría:

$$Y_i = X_i\beta + \varepsilon_i = \beta_0 + \beta_1x_{1,i} + \beta_2x_{2,i} + \dots + \beta_k(x_{k,i}^* - \vartheta_i) + u_i$$

Se estimaría la regresión:

$$Y_i = X_i\beta + \varepsilon_i = \beta_0 + \beta_1x_{1,i} + \beta_2x_{2,i} + \dots + \beta_kx_{k,i}^* + (u_i - \beta_k\vartheta_i)$$

Dado que $x_{k,i}^*$ es una construcción lineal de ϑ_i , claramente se encuentra la relación $cov(x_{k,i}^*, \vartheta_i) \neq 0$, lo cual representa una correlación entre la variable independiente $x_{k,i}^*$ y el término de error. De tal forma que la regresión es endógena, y el coeficiente β_k de la última regresión no reflejará el verdadero efecto de $x_{k,i}^*$ frente a Y_i . El error de medición en una de las variables independientes no es solucionable. Este tipo de problema solamente se puede conocer si se dispone de información sobre el levantamiento de información usada; y, dado que es no solucionable, la regresión no podrá reflejar el verdadero efecto causal de la variable con error de medición.

Simultaneidad

La simultaneidad se da cuando al menos una de las variables independientes $x_{j,i}$ es al mismo tiempo una de las causas de Y_i como un efecto. Un ejemplo de este tipo de endogeneidad es la tasa de criminalidad y la cantidad de policías en una ciudad, si se considera un modelo de regresión con la tasa de criminalidad como Y_i y la cantidad de policías como una $x_{j,i}$, debe tomarse en cuenta que la tasa de criminalidad se encuentra en cierta forma afectada por la cantidad de policías, pero al mismo tiempo la cantidad de policías se encuentra parcialmente determinada por la tasa de criminalidad (Wooldrige, 2002). La solución usual a este tipo de endogeneidad se da mediante modelos de ecuaciones simultáneas o variables instrumentales.

Se han analizado las tres causas generales de endogeneidad, permitiendo comprender la importancia de un modelo correctamente especificado y los detalles de las situaciones bajo las cuales se incumple la exogeneidad evitando el análisis causal de una regresión. A continuación se explica el análisis causal una vez que se logra garantizar la exogeneidad.

Análisis Causal

Una regresión será causal si la Función de Esperanza Condicional (FEC) lo es (Angrist & Pischke, 2009). Según lo analizado en la sección anterior se puede afirmar que la FEC será causal si no se da alguna de las tres situaciones que incumplen la exogeneidad. Cuando se quiere estudiar el efecto causal de un conjunto X_i de variables y estas no pueden garantizar un modelo exógeno se incluye un conjunto C_i de variables de control de tal forma que el modelo sea exógeno:

$$Y_i = X_i\beta + C_i\gamma + u_i$$

Tras lograr especificar correctamente el modelo se puede entender la relación causal bajo la idea de un contrafactual. Asumiendo que la variable sobre la cual se desea conocer el efecto causal es una variable dicotómica: $T=1$ o $T=0$. El efecto causal de T únicamente se conocerá con precisión si se pudiese observar al individuo i bajo $T=1$ simultáneamente que bajo $T=0$. Ya que es imposible observar a un individuo i en dos estados de la naturaleza, la mejor forma de estimar el efecto causal es mediante la esperanza condicional en una muestra lo suficientemente grande:

$$E[Y_i|T_i = 1] - E[Y_i|T_i = 0]$$

Por ejemplo, si T se refiere al internamiento en un hospital en los últimos doce meses, y Y_i el estado de salud, la diferencia de promedios no reflejará el efecto de la hospitalización en la salud, ya que las personas que se hospitalizaron presentaran un estado de salud menor frente a las personas que no. Lo que se desea conocer es el estado de salud promedio de las personas que se hospitalizaron si no lo hubiesen hecho:

$$E[Y_{i,1} - Y_{i,0}|T_i = 1]$$

Ya que no se puede observar al mismo individuo en dos estados de la naturaleza, y a la vez se requiere una aproximación a esta situación ideal, se plantea:

$$E[Y_i|C_i, T_i = 1] - E[Y_i|C_i, T_i = 0]$$

FEC que deberá garantizar exogeneidad. La anterior ecuación se interpreta como el efecto promedio de T dado el conjunto de variables de control C_i , de tal forma que se conoce en promedio el efecto del tratamiento T . El análisis de dicho efecto se da en el contexto *Ceteris Paribus*, donde se conoce el efecto promedio de T sobre Y , al mantener todos los demás factores constantes, a es. La generalización cuando el efecto de interés proviene de una variable continua o un conjunto de variables, se da en el caso más sencillo, mediante una regresión de Mínimos Cuadrados Ordinarios: $E[Y_i|X_i] = X_i\beta + \varepsilon_i$ (Angrist & Pischke, 2009). El efecto promedio de una variable $x_{k,i}$ ya sea continua o dicotómica se denomina el efecto parcial (Wooldridge, 2002). Para aproximar el efecto parcial de $x_{k,i}$ sobre Y_i manteniendo el resto de variables constantes, se efectúa la derivada parcial:

$$\frac{\delta E[Y_i|X_i]}{\delta x_{k,i}} = \beta_k$$

De tal forma que asumiendo un modelo correctamente especificado, β_k será el efecto parcial aproximado de $x_{k,i}$, pudiendo interpretarse como la variabilidad promedio que presentará Y frente a un cambio marginal de x_k , manteniendo el resto de variables fijas. Así β_k representará el efecto causal aproximado promedio. Como se presentó anteriormente un modelo correctamente especificado asume $E[\varepsilon_i|X_i] = 0$, supuesto denominado la Suposición de Independencia Condicional (SIC). Ya que no existe una comprobación estadística o matemática definitiva la SIC en un modelo de regresión, la demostración provendrá de la argumentación teórica de la existencia de la SIC dadas las variables escogidas para el modelo (Angrist & Pischke, 2009)

(Culyer y Newhouse, 2000) (Greene, 2010) (Wooldrige, 2002). Razón por la cual a continuación se presenta la argumentación teórica de las variables que deberían encontrarse en el modelo.

Para estimar $E(Y|X)$ se pueden tomar varios enfoques. Usar un modelo lineal de probabilidad $E(Y|X) = XB + \epsilon_i$, estimado mediante mínimos cuadrados ordinarios. O un modelo no lineal de probabilidad $E(Y|X) = P(Y_i=1|X_i)=F(X_iB)$, estimado mediante máxima verosimilitud, donde F puede ser una función de distribución acumulada normal o logística. Modelos recomendados para la estimación de variables como el acceso a servicios de salud (Culyer y Newhouse, 2000). Una vez generado el modelo de determinantes, la predicción en áreas pequeñas viene dada por la metodología descrita a continuación.

Salud y economía

El estado de salud de los individuos se encuentra explicado tanto por cuestiones médicas como por aspectos socio-económicos. Se encuentra que la prevalencia de enfermedades como la tuberculosis y las infecciones de transmisión sexual, se ven explicadas no solamente desde el punto de vista infeccioso, sino además, por factores sociales y económicos. Si se obviarán estos factores, los gobiernos comprenderían poco del fenómeno; y, por consiguiente, las decisiones tomadas no serán del todo exitosas. Además varios estudios confirman que los pobres tienden a enfermarse más, tanto en países ricos como en países pobres (Farmer, 1999).

Ya que la población pobre tiene menos acceso a la prevención y tratamiento efectivo de las enfermedades, tienden a tener más enfermedades y secuelas más graves, simultáneamente, se encuentran privados de los mejores servicios de salud, que se encuentran principalmente enfocados en la población más rica. Hay que tomar en cuenta que la propagación de enfermedades se encuentra determinada por elementos no económicos, por ejemplo, factores geográficos como la humedad y la temperatura ayudan a propagar enfermedades como la malaria, además de los factores económicos (Hamoudi & Sachs, 1999).

Numerosas enfermedades infecciosas requieren una respuesta inmunológica compleja, la cual se ve afectada por un estado de salud pobre y una mala alimentación, aumentando la gravedad de la enfermedad (Farmer, 1999). A nivel macro una mejoría de la salud de la población es tanto causa, como efecto del desarrollo económico de una nación.

Comprender los determinantes de la salud requiere una teoría multidisciplinaria donde se presente una alta colaboración entre ciencias médicas y ciencias sociales. Dentro de las ciencias médicas se comprende a la salud desde una óptica biológica, por otra parte, dentro de las ciencias sociales existen varios enfoques, desde los modelos cuantitativos en la economía al enfoque teórico de los antropólogos (Berman et al., 1994). En un intento por conjugar múltiples enfoques en una sola metodología Berman et al. proponen el concepto de Producción de Salud en el Hogar (PSH). Propuesta bajo la que se espera comprender los determinantes de la salud y sus cambios en el tiempo. PSH se enfoca en la presencia y mantenimiento de la salud, en lugar de estudiar la prevalencia de enfermedades. El estudiar enfermedades individualmente simplifica la complejidad de los efectos de la riqueza en un solo problema, lo que conlleva a la toma de decisiones erróneas. En PSH los hogares son la unidad principal de

análisis, sin significar que sean la única unidad de análisis, se incluyen también factores del entorno del hogar. Cabe recalcar que existe amplia experiencia en el fracaso de programas de salud que garantizan acceso a servicios, pero, estos no generan el impacto deseado, gracias al uso erróneo de los mismos (Berman et al., 1994). Bajo el enfoque de PSH los hospitales y clínicas dejan de ser los principales determinantes de la salud, ya que estos pasan a ser una de las opciones que tienen los hogares para mantener y mejorar su salud.

La conexión entre el comportamiento del hogar y la producción de salud, se da en base al entorno socio-cultural del hogar y a la vez por las decisiones del mismo. Según el enfoque PSH en primera instancia el hogar se ve afectado por el acceso de la comunidad a servicios de salud, agua limpia, contaminación medioambiental, alimentos, entre otras variables. En el mismo nivel se encuentran las características económicas del hogar, como educación e ingresos. Dentro de estas condiciones que acotan las posibilidades del hogar se dan comportamientos intra-hogar:

1. Prácticas de alimentación a los infantes del hogar: lactancia materna, tipo de comida, frecuencia, cantidad, calidad, etc.
2. Cuidado del niño: estimulación, cantidad y calidad de tiempo dedicado al niño, etc.
3. Comportamientos de búsqueda de salud como diagnóstico y tratamiento de enfermedades.
4. Comportamientos de limpieza, como costumbres de lavada de manos, almacenamiento de agua, mantenimiento de la limpieza del hogar con enfoque a evitar enfermedades, etc.
5. Cuidado pre y post natal de la mujer.
6. Uso de servicios de salud preventiva.
7. Uso de servicios curativos en base a calidad e intensidad.
8. Gasto en todo tipo de tratamiento.
9. Inversiones en bienes de capital relacionados con la salud

La *Nueva Economía del Hogar* (NEH) es un concepto desarrollado por Gary Becker de utilidad para comprender la demanda de salud, antes de la NEH la micro econometría analizaba a los fenómenos en dos esferas separadas: consumo y producción. En el marco conceptual de NEH los hogares compran o adquieren bienes con el fin de *producir* bienes de utilidad (commodities), mediante la combinación de bienes adquiridos, su tiempo y capacidad, siendo los bienes de utilidad el producto final deseado realmente por el hogar y la fuente de su satisfacción (Berman et al., 1994). Al igual que en la teoría microeconómica los agentes podrán generar múltiples combinaciones de bienes, tiempo y capacidades para obtener el mismo nivel de utilidad. Un ejemplo del enfoque NEH en relación a la salud es en la *producción* de niños sanos.

El enfoque NEH ha sido bastante criticado. Primero, los hogares pobres no tienen muchas posibilidades de elección. Además el modelo no toma en cuenta factores culturales y sociales. Por otro lado es un enfoque formal y rigurosamente desarrollado que permite estudiar independientemente los factores que afectan la producción de salud y el impacto de esta en el nivel de salud del hogar.

El análisis PSH abarca tanto el enfoque NEH como una aproximación adecuada a las restricciones a las que se enfrenta el hogar (primera esfera del análisis PSH). Hay que tener en cuenta que existen casos, donde, modelos micro econométricos no ayudan a comprender completamente los determinantes de un fenómeno. Por ejemplo, en un estudio antropológico dirigido por Carl Kendal se encuentra que, pese a una campaña gubernamental acompañada de visitas al hogar por trabajadores comunitarios, no se observa un aumento de la lactancia exclusiva hasta los 6 meses. Tras la recolección de información antropológica se descubre que los hombres de la comunidad consideraban, el tener relaciones sexuales con una mujer durante la lactancia arruinaría la leche materna. Por lo que el hombre intentaría que la mujer destete al bebe lo antes posible (Kendall, 1990).

Factores complejos como la espiritualidad, cultura, influencias sociales y económicas se encuentran contemplados en el análisis PSH, aunque muchos factores son difíciles de obtener mediante la información de censos y encuestas.

Demanda de Salud

El modelo desarrollado por Michael Grossman en 1972 es el estudio más importante en el ámbito de la demanda de salud, mediante el mismo se puede conocer en base a la teoría económica y el supuesto de racionalidad y maximización del beneficio del consumidor las variables que determinan la demanda de salud, construyendo un marco teórico mediante el cual se puede comprender cuestiones de discriminación, efectos del ingreso, la riqueza, la educación, entre otros. Dentro del modelo se asume los individuos heredan un stock de capital H_0 , este crecerá según:

$$H_t - H_{t-1} = I_{t-1} - \delta_{t-1}H_{t-1}$$

I_{t-1} será a la inversión en capital de salud en el periodo t-1, mientras que δ_{t-1} será la tasa de depreciación en t-1. δ será exógena y determinada por la edad del individuo. Se asumirá que la utilidad y el ingreso del individuo serán crecientes del capital de salud; y, que al decidir el valor óptimo de este capital, el individuo tomará en cuenta esta relación y el costo de mantener el capital en el momento de la optimización de utilidad. El stock de capital de equilibrio queda definido:

$$\tau_t + a_t = \{r + \delta_t - \hat{\pi}_{t-1}\}\pi_t$$

τ_t será el beneficio marginal pecuniario del capital en salud, a_t el beneficio no pecuniario, r la tasa de interés de, π_t es el costo marginal de inversión y $\hat{\pi}_{t-1}$ el costo marginal del cambio en el costo de inversión

Para poder estimar el modelo de forma empírica se deben generar formas funcionales de τ_t , r y π_t , asumiendo que $\delta_t - \hat{\pi}_{t-1}$ sea igual a 0 o alguna función del tiempo (Wagstaff A. , 2002). De tal forma que la ecuación estructural de demanda de atención médica quedará especificada:

$$\ln M_t = \ln H_t + (1 - \alpha_1) \ln w_t - (1 - \alpha_1) \alpha_1 \ln P_t^m + \alpha_2 t_t - \alpha_3 E$$

En lugar de estimar la ecuación anterior Grossman y Muurinen en 1982 estiman la forma reducida:

$$\ln M_t = [\alpha_1(\varepsilon - 1) + 1]\ln w_t - [\alpha_1(\varepsilon - 1) + 1]\ln P_t^m + \alpha_2(\varepsilon - 1)t_t + \alpha_3(\varepsilon - 1)E$$

En esta ecuación los coeficientes reflejaran tanto los parámetros de tecnología como la elasticidad de la demanda. Teóricamente la elasticidad ε debería ser menor a 1, por lo que en una comprobación empírica la educación debería tener signo negativo, comprobándose la hipótesis de la eficiencia de la educación. Manteniendo los demás factores constantes las personas más educadas deberían consumir menos servicios de salud, ya que estos son mejores productores de salud.

En el estudio empírico dirigido por Grossman y Muurinen en 1982 se obtienen resultados diferentes a los predichos por el modelo teórico, como por ejemplo un signo positivo de la variable de educación. Resultados que fueron justificados dado el supuesto de ajuste instantáneo de stock de capital y capital deseado. Según Wagstaff (2002) dicho supuesto no puede generar resultados empíricos tan diferentes al modelo teórico. Wagstaff plantea un modelo sin log-linearización, donde se partirá de:

$$H_t = \beta X_t + u_t$$

En dicho modelo se asume que el capital de salud (H_t), inversión en capital de salud (I_t) y demanda de servicios de salud (M_t) son variables latentes. Para estimar variables latentes, para su estimación Wagstaff utiliza modelos de Múltiples Indicadores Múltiples Causas (MIMIC), modelo que permite la estimación de la forma:

$$L = BX + e$$

Siendo L la variable latente a estimarse, X una columna de variables, B una columna de coeficientes y e una columna de errores. L es una variable no observable, la cual causa a cada variable de X más un término de error. La variable no observable es estimable bajo supuestos de normalidad multivariada; y, mediante máxima verosimilitud estimar los coeficientes o efecto marginal de una de las variables X sobre la variable latente. Además en el modelo de Wagstaff no se asume un ajuste instantáneo al nivel de capital de salud deseado, de tal forma que se especifica el capital de salud:

$$H_t - H_{t-1} = \mu(H_t^* - H_{t-1})$$

Donde H_t^* será el capital deseado, el modelo empírico de Grossman asume $\mu = 1$, supuesto considerablemente débil. Si bien el supuesto es aceptable en el caso de no disponer de data en varios periodos de tiempo. Wagstaff mediante información de la encuesta de salud Danesa para los años 1982 y 1983, obtiene resultados consistentes con el modelo teórico de Grossman, donde se argumenta una relación positiva entre la demanda de servicios de salud y la educación dada la linealidad inherente del modelo empírico planteado.

Planteamiento Empírico

En el Ecuador no existen encuestas de salud en la que se pueda realizar un análisis en varios periodos de tiempo, ya que la ENSANUT fue levantada en el año 2012 y la ENDEMAIN en el año 2004. Partiendo del modelo de Grossman se asumirá $\mu = 1$, planteando adquisición instantánea del capital de salud deseado para el periodo t , si bien el supuesto no es del todo real, es aceptable para data de corte transversal.

Se trabajará con una aproximación al planteamiento empírico del modelo de Wagstaff sobre el modelo teórico de Grossman. Aunque únicamente se trabajará con una variable latente: la demanda de servicios de salud. Se evitará plantear el resto de variables, como latentes ya que el objetivo de la disertación no es únicamente conocer el efecto de cada variable considerada en el modelo sobre la demanda de servicios de salud. Además se incluirá el modelo de Producción de Salud en el Hogar (PSH), ya que el modelo de Grossman no toma en cuenta características culturales, sociales y espirituales a las que se enfrenta el hogar.

El análisis de producción de salud del modelo PSH encaja correctamente con la representación empírica del modelo de Grossman planteada por Wagstaff. Ya que en los modelos MIMIC donde se caracterizaban las variables sobre inversión en capital de salud y stock de capital, se podría añadir información de la primera esfera de análisis de PSH, como los factores limitantes a los que se enfrenta el hogar (cultura, sociedad, economía y espiritualidad). Si bien intuitivamente se puede considerar esta compatibilidad entre estos modelo teóricos, escapa de los objetivos de la presente disertación una demostración formal de dicha afirmación.

Modelo de Grossman

Se considerará para la aplicación empírica la propuesta presentada por Michael Grossman en 1972, donde se utilizó la información de la Encuesta de Salud llevada por el Centro Nacional de Investigación de Opinión (*National Opinion Research Center* en inglés), que presenta la información de 2,367 familias y 7,803 personas con un muestreo probabilístico (Andersen & Anderson, 1968). Cabe recalcar que la información de dicha base de datos no se encuentra disponible al público en general, por lo que no se puede tener la certeza de replicar los métodos de utilizados por Grossman (Interuniversity Consortium for Political and Social Research).

Grossman estima la función de producción de salud partiendo de:

$$h_i = 365 - BH_i^C$$

Donde h_i será la función de producción de días saludable del individuo i , H_i la cantidad de días que la persona paso enferma en el año; y, B y C serán constantes positivas. La Eficiencia Marginal del Capital de Salud (MEC por sus siglas en inglés) según Grossman será:

$$\ln \gamma_i = \ln BC - (C + 1)\ln H_i + \ln W_i - \ln \pi_i$$

π_i será la tasa de depreciación del capital, que se encontrará asociada a la educación E y riqueza R acumulada de la persona. W_i serán los ingresos; y, dado que se asumen condiciones de equilibrio general la MEC será igual a la tasa de depreciación por lo que $\gamma_i = \delta_i$. Además la tasa de depreciación se expresará:

$$\ln \delta_i = \ln \delta_0 + \delta_i$$

A la vez se puede apreciar que el tiempo que la persona pase enferma (TL_i) se encontrará positivamente correlacionada con H_i por lo que el planteamiento empírico puede ser formulado:

$$h_i = 365 - TL_i$$

Por lo tanto:

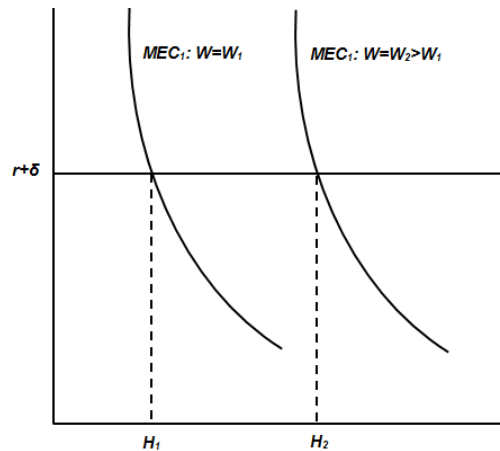
$$-\ln TL_i = -\ln B_i + C \ln H_i$$

Y a partir de las formulaciones de Grossman se tendrá que:

$$-\ln TL_i = CB_R \ln R + C(1 - K)\varepsilon \ln W_i + Cr_H \varepsilon E - C\delta \varepsilon i - C\varepsilon \ln \delta_0$$

Donde ε se refiere a la elasticidad de la eficiencia marginal del capital, y se espera que $0 < \varepsilon < 1$; K será el porcentaje del capital en salud deseado por el individuo que se comprará en el periodo estudiado, por lo que un aumento de 1% en los ingresos del individuo aumentaran en K por ciento el coste marginal del capital en salud. Ya que la demanda de salud se encuentra correlacionada positivamente con la demanda de salud, un aumento en los ingresos W de W_1 a W_2 aumentará la MEC como se puede apreciar en la Ilustración No. 1:

Ilustración No. 1



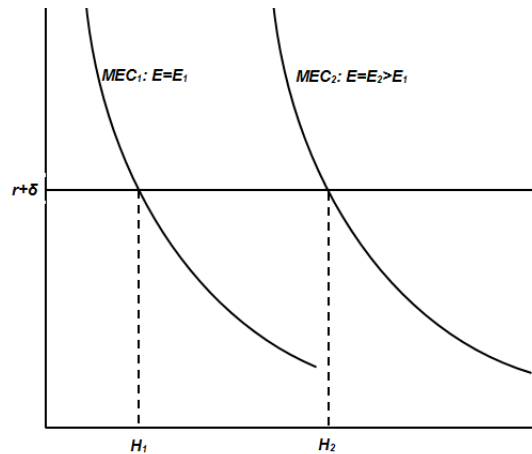
Fuente: Grossman, 1972

Elaboración: Diego Villacreses

Donde r_H se refiere al efecto que tendrá una variación de una unidad en educación E en relación a la inversión en salud (referencia al modelo de inversión pura de Grossman que en la

presente disertación no se desarrolla a mayor a profundidad). Ya que la educación aumenta la productividad se tiene que $r_H > 0$. La educación aumenta la capacidad de la persona de producir capital de salud, por lo que se necesitará una menor cantidad de *inputs* para producir un mismo nivel de capital en salud, dicha afirmación asume que los precios de los *inputs* se mantienen constantes al aumentar E . Manteniendo el resto de factores constantes, un aumento en la educación aumentará la Eficiencia Marginal del Capital, dicho razonamiento se grafica en la Ilustración No. 2.

Ilustración No. 2



Fuente: Grossman, 1972

Elaboración: Diego Villacreses

Pese a la importancia del Modelo de Grossman para la Teoría Económica y para la Economía de la Salud se han realizado pocos estudios empíricos con el fin de comprobar las hipótesis del modelo (Wagstaff A. , The demand for health: an empirical reformulation of the Grossman model, 2002). Los intentos de comprobar el modelo mediante el uso de los modelos estructurales propuestos por Grossman encontraron que empíricamente se rechaza el modelo (Wagstaff A. 1986) (Grossman, 1972).

Se deduce que la educación, los ingresos y la riqueza deben tener retornos positivos en relación a la demanda de salud. El término $C\delta\epsilon_i$ será una perturbación aleatoria que se añadirá al término de error sin perjudicar el supuesto de exogeneidad y $C\epsilon\ln\delta_0$ se añadirá a la constante.

Con el fin de comprobar y estimar la teoría de Grossman para el Ecuador se recurre a la Encuesta Nacional de Salud y Nutrición (ENSANUT) 2011-2013, la cual fue levantada entre julio y agosto del año 2012, se realizaron 12 formularios. Los cuales se detallan a continuación.

- Formulario Hogares: se entrevistaron a todos los miembros del hogar siendo un total de 92,502 personas en 19,949 hogares.
- Formulario Mujeres en Edad Fértil (MEF): se escogió a una mujer entre los 12 y 49 años de edad escogida aleatoriamente, se entrevistaron 18,213 mujeres en 18,213 hogares.

- Formulario de Lactancia Materna a Menores de 3 Años: se entrevista a la madre o tutor, se recoge información de todos los niños menores a 3 años del hogar. Se presentan 6,034 observaciones en 5,831 hogares.
- Formulario de Salud en la Niñez. se entrevista a las mujeres seleccionadas para el formulario MEF que tengan al menos un hijo menor a 5 años, recogiendo la información de todos sus hijos en dicho rango etario, con un total de 10,216 niños en 8,200 hogares.
- Formulario de Actividad Física: se escogen aleatoriamente entre 1 a 7 personas en el hogar que se encuentren entre los 18 y 60 años, se entrevistaron a 19,883 personas en 11,937 hogares.
- Formulario de Salud Sexual y Reproductiva en Hombres: la base de datos no se encuentra disponible al público. (Instituto Nacional de Estadísticas y Censos, 2014b)
- Formulario de Antropometría: se escoge aleatoriamente entre 1 a 11 personas menores a 60 años en el hogar. Se presenta información de 60,629 personas de 19,836 hogares.
- Formulario de Consumo: se entrevista a la persona encargada de preparar los alimentos en el hogar. Se levanta información en 10,008 hogares, entrevistándose a un total de 21,249 personas.
- Formulario de Bioquímica: Se escogen aleatoriamente entre 1 a 7 personas en el hogar menores a 60 años, se entrevistaron a 21,479 personas en 10,019 hogares.

Ya que en la presente disertación se estima un modelo estructural a partir de la formulación de Grossman y además se presentan diversos modelos de determinantes pudiendo conocer más relaciones entre variables observables y la demanda de salud que los predichos por Grossman se busca utilizar la mayor cantidad de información posible.

Se utilizará la ENSANUT que es abundante en información relacionada con la salud de las personas, a continuación se detalla en líneas generales la información presente en cada formulario. En el Formulario de Hogares se recoge información de la vivienda y el hogar, migración internacional, características generales, educacionales y socio-económicas, acceso a seguros de salud, uso de servicios y gasto en salud, acceso a programas sociales, seguridad alimentaria; y, mortalidad en el hogar. En el Formulario MEF se encuentra información de salud sexual y reproductiva, historial de embarazos, servicios de salud materna, conocimiento y uso de planificación familiar, preferencias reproductivas, actividad sexual y conocimiento de enfermedades de transmisión sexual.

En el Formulario de Lactancia Materna se presentan preguntas relacionadas con la lactancia temprana, lactancia exclusiva y continua; y, alimentación complementaria. Por otra parte el Formulario de Salud en la Niñez presenta información de atención en el parto, atención médica del recién nacido, control postparto, presencia de enfermedades diarreicas y respiratorias; y, vacunación. El Formulario de Actividad Física se recoge preguntas relacionadas al tiempo dedicado a la semana a caminar, correr, recreación y deporte. En el Formulario de Antropometría se encuentra información de peso, talla, cintura y presión de la persona. El Formulario de Consumo se refiere a un recordatorio del consumo de alimentos durante el día anterior a la entrevista con sus respectivos cálculos de componentes macro y micro nutrientes para la salud. Finalmente se encuentra que el Formulario de Bioquímica tiene información sobre

vitaminas, minerales, glucosa entre otros componentes hallados en la sangre de los encuestados.

Para el Modelo de Grossman se utilizará el Formulario de Hogares pues presenta toda la información necesaria: ingresos, educación, edad, sexo, afiliación al seguro social, riqueza y días que paso enfermo. Adicionalmente se realizaran varias versiones del modelo para comprobar la robustez de los resultados, se presentan 7 versiones del mismo, recolectadas en dos grupos. El primero maneja siempre toda la información presente en la muestra:

1. Se utiliza la escolaridad del jefe de hogar, el logaritmo natural del ingreso per cápita y los quintiles de riqueza para toda la muestra (los mismos se calculan a partir del índice proxy de riqueza)
2. Se utiliza la escolaridad del jefe de hogar, el logaritmo natural del ingreso per cápita y el índice proxy de riqueza, que se refiere a un variable latente construida a partir de características y equipamiento de la vivienda con el que se aproxima la riqueza (Freire , y otros, 2013). Se utiliza toda la muestra.
3. Se utiliza la escolaridad del jefe de hogar, el logaritmo natural del ingreso per cápita y el logaritmo del índice proxy de riqueza, el cual por ser una variable latente obtenida a partir de un análisis factorial originalmente presentaba promedio 0, por lo que se sumó el valor absoluto del valor mínimo de la distribución a toda la distribución, logrando así que la variable sea mayor a 0, a la cual se aplica el logaritmo natural.

En el segundo los 2 primeros se acotan a personas entre 18 y 65 años, los dos siguientes se encuentran acotados para personas entre 18 y 65 años y que se encuentren afiliadas al seguro social, esta última se presenta ya que en el desarrollo del modelo estructural se menciona que uno de los supuestos es que los precios son constantes entre los individuos, intentando contralar esta variable únicamente por el ingreso y la riqueza (Grossman, 1972), al considerar únicamente afiliados este supuesto se fortalece.

En la Tabla No. 4 y Tabla No. 5 se presentan los resultados del Modelo de Grossman aplicado a la información del Ecuador para el año 2012 en base a la ENSANUT, que como se puede apreciar se comprueban los resultados obtenidos por el autor en 1972 para el caso ecuatoriano en todas las versiones expuestas con anterioridad, comprobando así la robustez de los resultados.

Cabe recalcar una cuestión metodológica importante, ya que la ENSANUT es una muestra compleja, multietápica y estratificada (Freire , y otros, 2013); los errores estándares a usarse para obtener una correcta estimación deben ser los linealizados de Taylor corregidos por la cantidad Unidades Primarias de Muestreo y el Dominio Estadístico (Cameron & Trivedi, 2009) (Lohr, 2009). Este tipo de errores son computacionalmente difíciles de obtener por lo que no todos los comandos de STATA permiten su estimación, razón por la cual una aproximación altamente eficiente a los mismos y disponibles en casi todos los comandos de STATA son los errores clusterizados, además a medida que crece el tamaño muestral los dos errores estándares deberán converger, razones por los que se puede utilizar a los dos tipos de errores indiferentemente (Cameron & Trivedi, 2009). En la presente disertación se decidió utilizar en

todas las regresiones errores clusterizados por su facilidad de cálculo y flexibilidad de comparación entre diferentes modelos que de otra forma no hubiese sido posible.

En la ENSANUT no se recoge la información de enfermedades durante todo el año como en las muestras analizadas por Wagstaff en 1986 y Grossman en 1972, la base presenta únicamente información del último mes, la cual se considera más cercana a la realidad por el problema de recordatorio. La especificación del tiempo en el que la persona se encuentre saludable se formula:

$$h_i = 365 - TL_i$$

Se tendrá como variable dependiente $-\ln(TL_i)$, el modelo quedará especificado:

$$-\ln(TL_i) = \beta_0 + \beta_1 E + \beta_2 \ln W + \beta_3 \ln R + \beta_4 \text{Edad} + \beta_5 \text{Sexo} + e_i$$

Como se mencionó anteriormente $\ln W$ presentará especificaciones como ingresos per cápita o ingresos totales de la persona; y, $\ln R$ se aproximará mediante la Proxy de la Riqueza construida por Freire et al. en el 2013. La Tabla No. 4 presenta los resultados del primer grupo de regresiones, se encuentra que en las tres regresiones la escolaridad presenta signo negativo y el coeficiente es significativo al 99% de confianza. En la primera regresión un año adicional de escolaridad disminuye el tiempo saludable de la persona en 4.1 puntos porcentuales $(\exp(\beta_1) - 1) * 100$, en la segunda un año adicional representará 3.7 puntos porcentuales menos; y, en la última 2.7% puntos porcentuales. Estos valores no deben ser tomados como coeficientes bajos, ya que, en la primera regresión, el cursar una carrera universitaria de 5 años disminuiría la cantidad de días enfermo en 20.5 puntos porcentuales.

El Modelo Estructural expresaba claramente que el coeficiente de la Educación debía ser positivo y significativo, al igual que los estudios llevados por Grossman en 1972 y Wagstaff en 1986 se encuentra que la investigación empírica rechaza dicho modelo. Lo cual según una posterior investigación de Wagstaff en el 2002 no significa que el planteamiento teórico de la Demanda de Salud del autor haya sido incorrecta sino su formulación empírica. Por lo cual propone el uso de modelos de máxima verosimilitud que permitan el uso de ecuaciones recursivas al tener en cuenta H_{it} y H_{it-1} . Ya que las encuestas levantas en el Ecuador no permiten un análisis de ese tipo el desarrollo empírico de este modelo se dejará en el análisis de robustez.

Siguiendo con la Tabla No. 4 se aprecia que en las dos primeras especificaciones el \ln del ingreso per cápita no es significativo pero es positivo, lo cual es predicho por Grossman cuando el ingreso y la riqueza se encuentran altamente correlacionados, relación presentada en las dos primeras regresiones. En la primera regresión se aprecian que los quintiles 2 al 3 de riqueza no presentan un coeficiente significativo en relación a la categoría base, si bien los quintiles del 3 al 5 presentan coeficientes significativos, nótese que pasar del quintil 1 al 5 de riqueza aumenta el tiempo saludable de la persona en 189%.

La edad presenta signo negativo como predice el modelo teórico, a mayor edad aumentan la prevalencia e intensidad de enfermedades. Además se considera que las mujeres tienen menor

acceso a servicios de salud a lo largo de su vida por lo que se espera un coeficiente significativo y positivo (asumiendo que hombre =1) lo cual se observa en el Ecuador. Según los tres modelos los hombres en relación a las mujeres tienen alrededor de 91 puntos porcentuales más de tiempo saludable.

Nótese que las regresiones presentan 86,377 observaciones, ya que solamente 90,072 personas respondieron la información de enfermedad aunque se mantienen los 19,949 hogares. Por otro lado únicamente 18,918 hogares respondieron la información de ingresos, a los cuales pertenecen 86,377 personas. La medida de ajuste del modelo de regresión lineal R-cuadrado es alrededor de 0.01 lo cual no representa un problema dado el proceso generador de datos de la variable dependiente (Wagstaff A. , 2002) (Grossman, 1972).

Tabla No. 4
Estimaciones del Modelo de Grossman para toda la muestra. Variable dependiente: menos logaritmo natural de número de días enfermo.

VARIABLES	(1)	(2)	(3)
Escolaridad Jefe de Hogar	-0.0419*** (0.0104)	-0.0384*** (0.0105)	-0.0277*** (0.0103)
Ln de Ingreso per cápita del hogar	0.0551 (0.0549)	0.0515 (0.0554)	0.113** (0.0540)
Quintiles de Riqueza (base = 1)			
Q2	-0.0310 (0.115)		
Q3	0.0244 (0.116)		
Q4	0.435*** (0.137)		
Q5	1.062*** (0.151)		
Edad	-0.0216*** (0.00167)	-0.0216*** (0.00167)	-0.0211*** (0.00167)
Sexo (1=hombre)	0.650*** (0.0468)	0.654*** (0.0467)	0.650*** (0.0467)
Índice Proxy de Riqueza		0.363*** (0.0523)	
Ln de Índice de Proxy de Riqueza			0.624*** (0.128)
Constante	4.553*** (0.246)	4.839*** (0.270)	3.802*** (0.225)
Observaciones	86,377	86,377	86,377
R-cuadrado	0.013	0.011	0.010
Errores estándares Clusterizados en paréntesis			
*** p<0.01, ** p<0.05, * p<0.1			

Fuente: ENSANUT 2011-2013

Elaboración: Diego Villacreses

En la Tabla No. 5 se presentan cuatro regresiones y todas se encuentran condicionadas al grupo etario entre 18 y 65 años de edad, ya que las bases de datos que utilizaron tanto Grossman como Wagstaff se encontraban limitadas a personas mayores de 18 años por lo que se decide presentar la Tabla No. 5. En las dos primeras regresiones se corroboran los

resultados presentados en la Tabla No. 4, con coeficientes negativos y significativos para la escolaridad del jefe de hogar y la escolaridad del individuo. En la primera regresión el logaritmo del ingreso per cápita es no significativo lo cual se puede deber a la alta correlación que presenta con la riqueza (0.5653).

La edad tiene un coeficiente casi igual en las cuatro regresiones, alrededor de -0.064, lo cual lleva a las mismas conclusiones que las regresiones presentadas en la sección anterior. El Índice Proxy de Riqueza es positivo y significativo, además presenta ligeras variaciones en las cuatro especificaciones presentadas. Al analizar únicamente personas entre 18 y 65 años se encuentra que la diferencia entre hombres y mujeres es más acentuada, ya que los hombres tienen un 197% más de tiempo saludable en relación a las mujeres, situación que se refleja de forma similar en las cuatro especificaciones.

Puede notarse que en las regresiones 3 y 4 que se trabaja únicamente con personas afiliadas al seguro social, la escolaridad de la persona o del jefe de hogar no son significativas, lo cual no corresponde con el Modelo Teórico. Fenómeno que se puede deber a la disminución de observaciones o a una fuerte relación entre la afiliación al seguro social y la educación.

Tabla No. 5

Estimaciones del Modelo de Grossman para personas entre 18 y 65 años; y, personas en dicho grupo que tienen seguro social. Variable dependiente menos logaritmo natural de número de días enfermo.

VARIABLES	(1)	(2)	(3)	(4)
Escolaridad Jefe de Hogar	-0.0252** (0.0111)			-0.00849 (0.0167)
Ln de Ingreso per cápita del hogar	0.0469 (0.0644)			0.143 (0.108)
Índice Proxy de Riqueza	0.365*** (0.0597)	0.279*** (0.0611)	0.341*** (0.0906)	0.409*** (0.0954)
Edad	-0.0643*** (0.00311)	-0.0635*** (0.00431)	-0.0645*** (0.00668)	-0.0625*** (0.00555)
Sexo (1=hombre)	1.034*** (0.0628)	1.093*** (0.0978)	1.111*** (0.154)	1.234*** (0.125)
Escolaridad		-0.0340** (0.0139)	-0.0133 (0.0210)	
Ln del Ingreso		0.227*** (0.0701)	0.309** (0.131)	
Constante	6.349*** (0.326)	5.292*** (0.395)	4.430*** (0.727)	5.324*** (0.590)
Observaciones	45,801	29,178	12,059	14,761
R-cuadrado	0.031	0.028	0.029	0.034
Errores estándares Clusterizados en paréntesis				
*** p<0.01, ** p<0.05, * p<0.1				

Fuente: ENSANUT 2011-2013

Elaboración: Diego Villacreses

La presente sección analizará bajo la misma lógica de Grossman, pero considerando variables de control no utilizadas en estudios anteriores. Las mujeres presentan 6 puntos porcentuales más de prevalencia de enfermedades (diferencia que es estadísticamente significativa), y 1 día más en promedio de tiempo de enfermedad. Por lo que cabe resaltar que en la ENSANUT se recoge información sobre el tipo de enfermedad que tuvo el informante, las cuales se detallan a continuación:

- Respiratorios: Gripe común, sinusitis, rinosinusitis, faringitis/adenitis, laringitis, traqueitis, bronquitis, bronquiolitis, neumonía, otitis media, tuberculosis, asma, enfisema. Los síntomas principales incluyen congestión nasal, tos, rinitis, dolor de garganta, fiebre y estornudos.
- Cardiovascular: Infartos, derrames, insuficiencia cardíaca, soplos al corazón, angina, aterosclerosis, accidente cerebro vascular apoplejía, enfermedades cerebro vasculares, enfermedad de la arteria coronaria, enfermedad vascular periférica, várices, trombosis.
- Digestivos: Enfermedad diarreica aguda, parasitosis, infecciones digestivas, salmonelosis, fiebre tifoidea, intoxicación alimentaria, cólera, indigestión, apendicitis, gastritis, colitis, úlceras pépticas, hepatitis, fibrosis quística, colecistitis (infección en la vesícula biliar), colelitiasis (cálculos en la vesícula biliar), pancreatitis.
- Problemas de la piel: Dermatitis, roncha, acné, psoriasis, sarpullido, eccema, hiperpigmentación, rosácea, pecas, herpes, verrugas.
- Problemas de los ojos: Desprendimiento de la retina, hipermetropía, estrabismo, miopía, glaucoma, cataratas, orzuelos, degeneración macular
- Problemas neuromusculares: Mialgias (dolor muscular), neuritis, distrofia muscular (debilidad muscular y pérdida de la masa muscular), torceduras y distensiones, calambres o tendinitis, miositis (inflamación).
- Problemas dentales: Caries, gingivitis, piorrea, dolor de muelas, dientes impactados, bruxismo.
- Problemas del embarazo: fatiga, acidez, problemas de espalda, cambios de humor, trastornos del sueño, náuseas matutinas, sofocos y sudores, calambres en las piernas, estreñimiento.
- Problemas psicológicos: Depresión, ansiedad, fobias, adicciones, obsesiones, inseguridad, timidez, estados de ánimo cambiantes, hipocondría (miedo a padecer una enfermedad), trastornos del sueño, en general comportamientos inadecuados que impiden a las personas adaptarse a su medio.
- Problemas óseos: artrosis, osteoporosis, enfermedad de Paget, osteoartritis, escoliosis, lordosis.
- Fracturas, heridas, golpes
- Enfermedades crónicas: Hipertensión arterial, diabetes, obesidad, cáncer, artritis otros tipos de cáncer.
- Enfermedad de la mujer: Falta de menstruación (no por causas del embarazo), menstruaciones irregulares, menstruaciones dolorosas, quistes en ovarios, mastitis, menopausia, tumor en las mamas, dolor permanente en la pelvis, infecciones vaginales,

infertilidad, inflamación de los ovarios, inflamación de las trompas de Falopio, incontinencia, cáncer de útero, mama, vagina, ovario.

- Enfermedad del hombre: Prostatitis, varicocele, enfermedades del testículo, inflamaciones del glande, cáncer del pene, testículos, próstata, impotencia

La hipótesis de inequidad en la presencia de Stock de Capital de Salud sugiere que las mujeres reciben menores flujos de capital de salud durante la infancia y adolescencia; y, durante la adultez tienen menos oportunidades de acceso a servicios de salud. Para contrastar dicha hipótesis correctamente se sugiere eliminar el tipo de enfermedades que se podrían considerar no relacionadas con los flujos de capital de salud descritos en el modelo. Para lo cual se propone analizar las regresiones sin considerar los problemas del embarazo y las enfermedades de la mujer; si bien algunas de las enfermedades de la mujer si se encuentran relacionadas con problemas en discriminación de flujos de capital es de interés comparar la prevalencia e intensidad de enfermedades no relacionadas con el género.

En la Tabla No. 6 se compara la prevalencia de enfermedades al tomar en cuenta todas las enfermedades, excluyendo problemas del embarazo, excluyendo enfermedades del hombre y la mujer y al excluir a las dos anteriores. Nótese como en las cuatro comparaciones a medida que aumenta el quintil económico disminuye la brecha entre géneros. Al tomar en cuenta todas las enfermedades, la prevalencia es de 45.7% para mujeres y 39.7% para hombres, como se puede apreciar los intervalos de confianza no se solapan, por lo que la diferencia estadísticamente significativa. En el primer quintil la diferencia es de 7.7 puntos porcentuales mientras que en el quintil 5 es de 5.7 puntos.

Al analizar las enfermedades excluyendo problemas asociados al embarazo, la prevalencia de problemas de salud en las mujeres baja del 45.7% al 45.6%, manteniéndose las brechas entre quintiles de forma similar. Al excluir las enfermedades de la mujer y el hombre, se aprecia que el 41.4% de las mujeres presenta alguna enfermedad frente al 39.1% de los hombres, lo que representa 2 puntos porcentuales de diferencia, nótese que la brecha en el quintil más bajo es de 4 puntos mientras que en el quintil más alto es de 0.5, diferencia que es estadísticamente insignificante. Finalmente al analizar la información en la que se excluye excluyendo problemas embarazo, enfermedades de la mujer y del hombre, se aprecia que se diferencia de la anterior por 0.1 o 0.2 puntos porcentuales por lo que las conclusiones no serán mayormente diferentes.

Tabla No. 6
Comparación de prevalencia de enfermedades

Quintil de Riqueza	Mujer		Hombre		Total	
	Promedio	IC	Promedio	IC	Promedio	IC
Todas las Enfermedades						
Q1	47.9	[46.2,49.6]	40.2	[38.5,41.9]	44.1	[42.6,45.5]
Q2	47.9	[46.2,49.6]	42.2	[40.4,43.9]	45.1	[43.6,46.6]
Q3	49.1	[47.5,50.8]	41.6	[39.9,43.3]	45.5	[44.1,46.9]
Q4	44.6	[42.6,46.5]	39.2	[37.3,41.1]	42	[40.3,43.7]
Q5	39.2	[37.1,41.3]	35.5	[33.5,37.5]	37.4	[35.6,39.2]
Total	45.7	[44.9,46.6]	39.7	[38.8,40.6]	42.8	[42.0,43.6]
Enfermedades excluyendo problemas del embarazo						
Q1	47.8	[46.1,49.5]	40.2	[38.5,41.9]	44	[42.6,45.4]
Q2	47.6	[45.9,49.3]	42.2	[40.4,43.9]	44.9	[43.5,46.4]
Q3	49	[47.3,50.6]	41.6	[39.9,43.3]	45.4	[44.0,46.8]
Q4	44.4	[42.4,46.3]	39.2	[37.3,41.1]	41.9	[40.2,43.5]
Q5	39.1	[37.1,41.2]	35.5	[33.5,37.5]	37.4	[35.6,39.1]
Total	45.6	[44.7,46.4]	39.7	[38.8,40.6]	42.7	[42.0,43.5]
Enfermedades excluyendo enfermedades de la mujer y del hombre						
Q1	43.5	[41.9,45.1]	39.5	[37.8,41.1]	41.5	[40.1,42.9]
Q2	43.3	[41.6,44.9]	41.6	[39.8,43.3]	42.4	[41.0,43.9]
Q3	44.1	[42.5,45.8]	41.1	[39.4,42.8]	42.7	[41.3,44.1]
Q4	40.7	[38.7,42.6]	38.5	[36.6,40.4]	39.6	[38.0,41.2]
Q5	35.6	[33.6,37.7]	35	[33.0,37.0]	35.3	[33.6,37.0]
Total	41.4	[40.6,42.3]	39.1	[38.2,40.0]	40.3	[39.5,41.1]
Enfermedades excluyendo problemas embarazo, enfermedades de la mujer y del hombre						
Q1	43.3	[41.7,44.9]	39.5	[37.8,41.1]	41.4	[40.0,42.8]
Q2	43	[41.4,44.6]	41.6	[39.8,43.3]	42.3	[40.8,43.7]
Q3	44	[42.4,45.6]	41.1	[39.4,42.8]	42.6	[41.2,44.0]
Q4	40.4	[38.5,42.3]	38.5	[36.6,40.4]	39.5	[37.9,41.1]
Q5	35.6	[33.5,37.6]	35	[33.0,37.0]	35.3	[33.6,37.0]
Total	41.3	[40.4,42.1]	39.1	[38.2,40.0]	40.2	[39.5,41.0]

Fuente: ENSANUT 2011-2013

Elaboración: Diego Villacreses

En la Tabla No. 7 se comparan los modelos de regresión por MCO en los que se cambia la variable dependiente. El primer modelo corresponde al logaritmo natural del total de días que la persona paso enferma en el último mes, mientras que en el segundo modelo se excluyen las enfermedades del hombre y la mujer con el fin de sobredimensionar la diferencia de tiempo de enfermedad entre género. Nótese la segunda regresión, se encuentra que el efecto de la escolaridad, el ingreso, la riqueza y la edad mantienen el mismo signo que en la regresión anterior pero los coeficientes son más pequeños. Por lo que se puede afirmar que tanto individualmente como en conjunto dichas variables disminuyeron su efecto al excluir del análisis enfermedades del hombre y la mujer, lo que se puede comprobar además con la disminución del r-cuadrado de 0.028 a 0.02.

El objetivo principal de realizar esta nueva regresión es presentar una estimación más certera del efecto que tiene el género en el tiempo que la persona se encuentra saludable. Nótese que el coeficiente pasó de 1.093 a 0.389, si bien el coeficiente aún es significativo este paso a ser casi un tercio que el anterior. Al incluir todo tipo de enfermedades los hombres mayores a 18

años tenían alrededor de 200 puntos porcentuales más en el tiempo que pasaban saludables, mientras que al excluir las enfermedades del hombre y la mujer, esta diferencia pasa a ser de 47 puntos porcentuales.

Tabla No. 7
Comparación de Modelos de Regresión

VARIABLES	(1)	(2)+
Escolaridad	-0.0340** (0.0139)	-0.0248** (0.0110)
Ln del Ingreso	0.227*** (0.0701)	0.150*** (0.0564)
Índice Proxy de Riqueza	0.279*** (0.0611)	0.180*** (0.0485)
Edad	-0.0635*** (0.00431)	-0.0504*** (0.00332)
Sexo	1.093*** (0.0978)	0.389*** (0.0809)
Constante	5.292*** (0.395)	4.512*** (0.321)
Observaciones	29,178	29,178
R-cuadrado	0.028	0.020

Errores estándares Clusterizados en paréntesis

*** p<0.01, ** p<0.05, * p<0.1

+ se excluyen enfermedades de la mujer y del hombre

Fuente: ENSANUT 2011-2013

Elaboración: Diego Villacreses

En esta sección se analizó realizó una aplicación empírica del Modelo de Grossman, llegando a las mismas conclusiones que el autor en 1972, el cual argumentó que la contradicción entre la aplicación empírica y el modelo teórico se podía deber principalmente al supuesto mediante el cual se elimina la intertemporalidad del modelo, pues teniendo que:

$$H_t - H_{t-1} = \mu(H_t^* - H_{t-1})$$

Grossman asume que μ es igual a 0, por lo que los individuos llegaran al nivel ideal de H_t^* para cada periodo en mismo periodo. Según Wagstaff en 1986, este supuesto no es el culpable del sesgo que presenta el coeficiente de la escolaridad, por lo que propone la construcción de un modelo de variables latentes que permita la inclusión de ecuaciones recursivas, como se mencionó anteriormente. Razón por la cual se plantearan modelos de regresión enmarcados dentro del Modelo Teórico, pero que no justifiquen con la riguridad del modelo teórico la presencia de todas sus variables independientes, teniendo como fin comprender la relación de la demanda de salud con un conjunto más amplio de variables. Las siguientes secciones presentan modelos que trabajen bajo dicha lógica.

Otros determinantes de la Demanda de Salud

Siguiendo la propuesta del modelo de la Producción de Salud en el Hogar de Berman, que fue expuesto ya en el capítulo teórico, se espera complementar el modelo de Grossman de demanda de salud ya que este únicamente analiza la salud en relación a los ingresos, riqueza, educación, sexo y edad. En base a la óptica del concepto PSH se pueden incluir variables comportamentales y socio-demográficas en el modelo.

Además el Modelo de Grossman limitaba el uso de la variable de consumo de servicios de salud únicamente en el caso de poseer información de precios del consumo, lo cual si existe en la ENSANUT, pero dicho análisis sería sumamente limitante en el caso ecuatoriano pues existe una fuerte presencia de servicios de salud gratuitos y uso de servicios de salud en base al seguro social por lo que su costo es igual a cero, como se expuso anteriormente por dicha razón se estimó únicamente la función de producción de días saludable, la cual es una función de la demanda de salud.

En base a dichas ideas se presentan las siguientes secciones, donde se analiza con mayor libertad la demanda de salud y la prevalencia de enfermedades. En la presente sección se analiza la demanda de salud, la prevalencia de enfermedades y la intensidad de las mismas.

Si la persona se enfermó o no es una variable dicotómica del tipo:

$$Y_i \begin{cases} 1 & \text{si la persona se enfermó en los ult. 30 días} \\ 0 & \text{caso contrario} \end{cases}$$

En la Tabla No. 8 se analizan 3 variables dependientes, la primera es igual a 1 si la persona se enfermó en los últimos 30 días. La segunda es 1 si la persona se enfermó y visitó un centro de salud o médico, se automedicó o se internó en un hospital, se excluye a las personas que no tomaron ninguna acción. La tercera excluye además automedicación. La primera puede considerarse como la demanda de salud más la demanda insatisfecha, la segunda como la demanda de salud y la última la demanda específica de servicios médicos.

La especificación de los modelos queda expresada de la siguiente forma:

$$P(Y_i = 1|X_i) = \text{logit}(\beta X_i)$$

Donde el conjunto de X incluirán las mismas variables utilizadas en la aplicación empírica del Modelo de Grossman. Además se incluirán variables de control como la etnia, el área de residencia, cantidad de miembros del hogar menores de 5 años y mayores de 65; y, variables que se podrían considerar causales, como el acceso a agua potable, servicios higiénicos, piso de tierra y cuarto exclusivo para cocinar.

Tabla No. 8
Regresión Logística de probabilidad enfermarse, demandar salud y demandar servicios de salud

Variables	(1)		(2)		(3)	
	Logit	dy/dx	Logit	dy/dx	Logit	dy/dx
Edad	0.0276*** (0.00737)	0.00644*** (0.00171)	0.0245*** (0.00530)	0.00513*** (0.00111)	0.0240*** (0.00657)	0.00338*** (0.000925)
Edad al cuadrado	-7.81e-05 (8.78e-05)	-1.82e-05 (2.05e-05)	-3.77e-05 (6.33e-05)	-7.89e-06 (1.33e-05)	4.17e-06 (7.70e-05)	5.88e-07 (1.08e-05)
Logaritmo de Ingreso Total Hogar	-0.105*** (0.0265)	-0.0246*** (0.00616)	-0.107*** (0.0192)	-0.0224*** (0.00402)	-0.0916*** (0.0229)	-0.0129*** (0.00324)
Índice de Proxy de Riqueza	-0.0665** (0.0316)	-0.0155** (0.00737)	-0.0342 (0.0242)	-0.00717 (0.00507)	-0.00851 (0.0291)	-0.00120 (0.00410)
Escolaridad Jefe de Hogar	0.00832* (0.00482)	0.00194* (0.00112)	0.0110*** (0.00374)	0.00230*** (0.000782)	0.0161*** (0.00437)	0.00227*** (0.000616)
Etnia (base = indígena)						
Afro/negro/mulato	0.531*** (0.105)	0.122*** (0.0241)	0.375*** (0.0897)	0.0769*** (0.0188)	0.131 (0.101)	0.0182 (0.0142)
Montubio	0.577*** (0.0983)	0.133*** (0.0226)	0.478*** (0.0853)	0.0998*** (0.0181)	0.195* (0.105)	0.0276* (0.0151)
Mestizo/otro	0.348*** (0.0702)	0.0784*** (0.0152)	0.263*** (0.0565)	0.0528*** (0.0109)	0.0898 (0.0680)	0.0123 (0.00913)
Blanco	0.583*** (0.140)	0.134*** (0.0329)	0.450*** (0.107)	0.0934*** (0.0231)	0.219* (0.129)	0.0312 (0.0191)
Área (1=Urbano)	0.217*** (0.0495)	0.0506*** (0.0115)	0.183*** (0.0403)	0.0384*** (0.00843)	-0.0430 (0.0472)	-0.00606 (0.00666)
Sexo (hombre =1)	-0.421*** (0.0363)	-0.0982*** (0.00833)	-0.423*** (0.0268)	-0.0888*** (0.00556)	-0.606*** (0.0322)	-0.0854*** (0.00454)
Cantidad personas 0-5 años	-0.0690*** (0.0249)	-0.0161*** (0.00581)	-0.104*** (0.0191)	-0.0217*** (0.00400)	-0.0963*** (0.0223)	-0.0136*** (0.00313)
Cantidad personas 65 más	-0.0479 (0.0436)	-0.0112 (0.0102)	-0.116*** (0.0327)	-0.0244*** (0.00684)	-0.104*** (0.0366)	-0.0146*** (0.00515)
Agua Potable	0.0157 (0.0536)	0.00366 (0.0125)	0.0561 (0.0446)	0.0118 (0.00935)	0.0629 (0.0539)	0.00887 (0.00760)
Servicios Higiénicos	-0.236*** (0.0477)	-0.0549*** (0.0111)	-0.171*** (0.0396)	-0.0358*** (0.00829)	-0.00465 (0.0467)	-0.000655 (0.00658)
Piso de Tierra	-0.0346 (0.0960)	-0.00806 (0.0224)	-0.0617 (0.0812)	-0.0129 (0.0170)	-0.131 (0.100)	-0.0185 (0.0142)
Cuarto exclusivo para cocinar	-0.182*** (0.0576)	-0.0425*** (0.0134)	-0.176*** (0.0430)	-0.0369*** (0.00899)	-0.121** (0.0536)	-0.0170** (0.00757)
Seguro Social	0.125*** (0.0389)	0.0292*** (0.00906)	0.135*** (0.0297)	0.0282*** (0.00623)	0.342*** (0.0358)	0.0481*** (0.00503)
Constante	-0.703*** (0.206)		-0.982*** (0.160)		-1.797*** (0.205)	
Observations	30,632	30,632	30,632	30,632	30,632	30,632
Robust standard errors in parentheses						
*** p<0.01, ** p<0.05, * p<0.1						

Fuente: ENSANUT 2011-2013
Elaboración: Diego Villacreses

Al igual que en el Modelo de Grossman se encuentra que al aumentar la educación disminuye el tiempo saludable de la persona, la primera regresión concluye que un aumento de un año de escolaridad aumenta la probabilidad de enfermarse en alrededor de 1 punto porcentual, y el coeficiente es estadísticamente significativo, lo cual contradice la teoría pese a considerar más variables de control. La siguiente regresión representa la demanda de salud, la cual según Grossman debería disminuir con la educación pues las personas más educadas son productores más eficientes de capital de salud, como se aprecia en la Tabla No. 8 dicha hipótesis no se confirma pues un año adicional de educación aumenta la demanda de salud o probabilidad de solicitar un servicio de salud en 1.1 puntos porcentuales.

Tanto para probabilidad de enfermarse, demanda de salud y demanda de servicios de salud los hombres presentan una menor probabilidad, alrededor de 8 puntos porcentuales, en relación a las mujeres. Se encuentra que de entre todas las etnias, los indígenas presentan una menor probabilidad tanto de enfermarse como de demandar salud. Mientras que los blancos son los más propensos a presentar enfermedades y por consiguiente demandar salud, en relación a los indígenas presentan 13.4 puntos porcentuales más de probabilidad de enfermarse y 26 puntos porcentuales más de demandar salud, al analizar la demanda de servicios de salud se encuentra que la etnia no influye significativamente.

Además se aprecia que el acceso a agua potable no afecta ni la probabilidad de enfermarse, ni la de demandar salud o servicios de salud, mientras que la presencia de servicios higiénicos en el hogar disminuye el 5.4 puntos porcentuales la probabilidad de enfermarse, en 3.5 puntos la probabilidad de demandar salud y es no significativa frente a la demanda de servicios de salud. La presencia de piso de tierra en la vivienda no afecta significativamente en ninguna de las tres regresiones.

Se encuentra que las personas del área urbana presentan una mayor probabilidad de presentar enfermedades y de demandar salud, presentando una diferencia de 5 puntos porcentuales en relación a las personas de las mismas características de enfermarse y 3.8 puntos en la demanda de salud.

La afiliación al seguro social representa 2.9 puntos porcentuales más a la probabilidad de enfermarse y 2.9 a la demandar salud y 4.8 a la demanda de servicios de salud, es contraintuitivo y probablemente endógeno un coeficiente positivo en la probabilidad de enfermarse, aunque probablemente se debe a los mismos problemas que generan un signo erróneo en la educación del jefe de hogar.

La presencia de un cuarto exclusivo para cocinar influencia de forma significativa a las variables analizadas. Disminuye la probabilidad de enfermarse en 4 puntos porcentuales, la de demandar salud en 1.8 y la de demandar servicios de salud en 0.1, si bien es un coeficiente considerablemente pequeño resulta ser significativo.

Con el fin de conocer la capacidad de predicción o clasificación para variables dicotómicas existen varias técnicas, en la presente disertación se discutirán las dos más utilizadas. En primer lugar se presenta la matriz de confusión, donde se encuentra la sensibilidad, la que representa la proporción de $Y=1$ correctamente clasificados por el modelo, la especificidad que

es la proporción de $Y=0$ correctamente predicho por el modelo, los falsos positivos que vienen a ser la proporción que realmente son $Y=0$ pero el modelo los consideró $Y=1$, los falsos negativos vienen a ser los $Y=1$ clasificados como $Y=0$, finalmente la clasificación global se referirá al total de valores correctamente clasificados, o la capacidad general de clasificación del modelo (STATACORP, 2014).

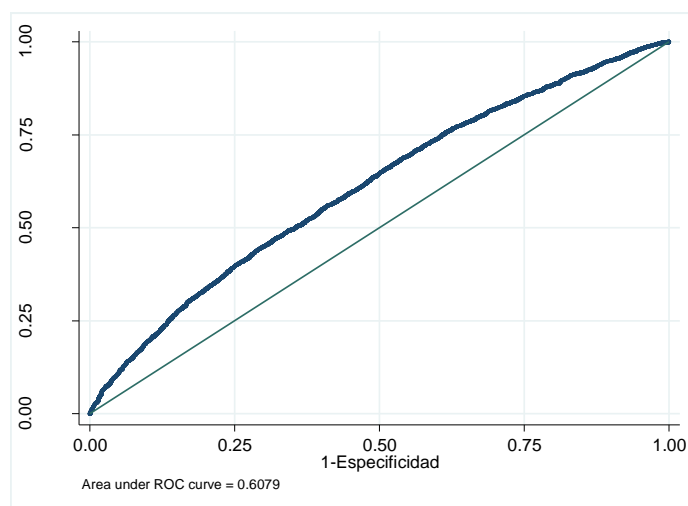
Cuando se caracteriza a un fenómeno Y_i mediante $P(Y_i = 1|X_i) = f(\beta X_i)$, en este caso f será una función logística, se puede realizar la predicción $\hat{Y}_i = f(\hat{\beta} X_i)$, donde $\hat{Y}_i \in [0,1]$, pero además $\hat{Y}_i \in \mathbb{R}$, por lo que la predicción podrá ser cualquier valor entre 0 y 1, con el fin de generar la clasificación se debe tener un punto de corte a partir del cual i será considerado 1; y, 0 en el caso de encontrarse por debajo. La matriz de confusión presentada por STATA originalmente asume un punto de corte 0.5 el cual no es adaptable a todas las distribuciones empíricas observables. Por ello se propone el uso de dos tipos de corte, el primero será el valor empírico de la proporción de $Y=0$, el segundo un punto donde maximice simultáneamente la sensibilidad como la especificidad, lo que es equivalente a la minimización de la diferencia entre las dos, lo cual puede ser comprendido como el punto de corte de las funciones teniendo como X los valores del punto de corte entre 0 y 1 (Cameron & Trivedi, 2009).

El primer punto de corte representa la aproximación más sencilla a una maximización de la predicción global, mientras que el segundo cuando por alguna razón se desee maximizar la proporción correctamente clasificada de 0 y 1 (Cameron & Trivedi, 2009), por ejemplo al desear maximizar las utilidades de un banco al desear dar crédito a la mayor cantidad de potenciales clientes y rechazar a los potenciales deudores asumiendo que todos requieran el mismo monto para su préstamo. Además se pueden realizar varias modificaciones a dicho enfoque con el fin de optimizar diferentes funciones objetivo.

La segunda técnica de clasificación es el uso de la curva *Receiver Operating Characteristic* (ROC) que permite conocer la clasificación promedio del modelo en relación a todos los posibles puntos de corte asignados a la matriz de clasificación, mediante un gráfico donde el eje Y será la sensibilidad y el X será 1-especificidad, evaluado para todos los puntos de corte entre 0 y 1 (STATACORP, 2014). En conjunto la curva ROC y la matriz de clasificación con los dos puntos de corte sugeridos pueden dar una idea general de la capacidad predictiva del modelo.

La Ilustración No. 3 es la Curva ROC de la segunda regresión de la Tabla No. 8, que corresponde a la regresión de demanda de salud, se puede apreciar que el área bajo la curva ROC es igual a 0.6067, lo cual significa que el modelo tiene una capacidad de predicción aceptable. Las curvas ROC para probabilidad de enfermarse y demanda de servicios de salud se encuentran en el Anexo 2, siendo la Ilustración No. 7 y la Ilustración No. 9 respectivamente, con valores de 0.6121 y 0.6252 en cada caso, lo que representa modelos aceptables.

Ilustración No. 3
Curva ROC de regresión logística sobre demanda de salud



Fuente: ENSANUT 2011-2013
Elaboración: Diego Villacreses

Nótese la Tabla No. 9 donde se presenta la matriz de confusión para la demanda de salud con un punto de corte igual a la proporción de casos que cumplen con $Y=1$ o lo que es equivalente $\Pr(Y=1)$, para la demanda de salud el punto de corte es 0.3392, se encuentra que se clasifican correctamente el 59.1% de los $Y=1$ y el 55.35% de los $Y=0$, además se encuentra que la clasificación global es igual a 56.6%. En el Anexo 2 la Tabla No. 18 corresponde a la matriz de confusión de la probabilidad de enfermarse, donde se encuentra que la sensibilidad es 59.1%, la especificidad 55.35% y la predicción global de 56.64% con un punto de corte de 0.3392. La Tabla No. 20 del Anexo 2 es la matriz de confusión de la demanda de servicios de salud con un punto de corte igual a 0.2114, sensibilidad de 47.94%, especificidad de 70.79% y predicción global de 66.53%.

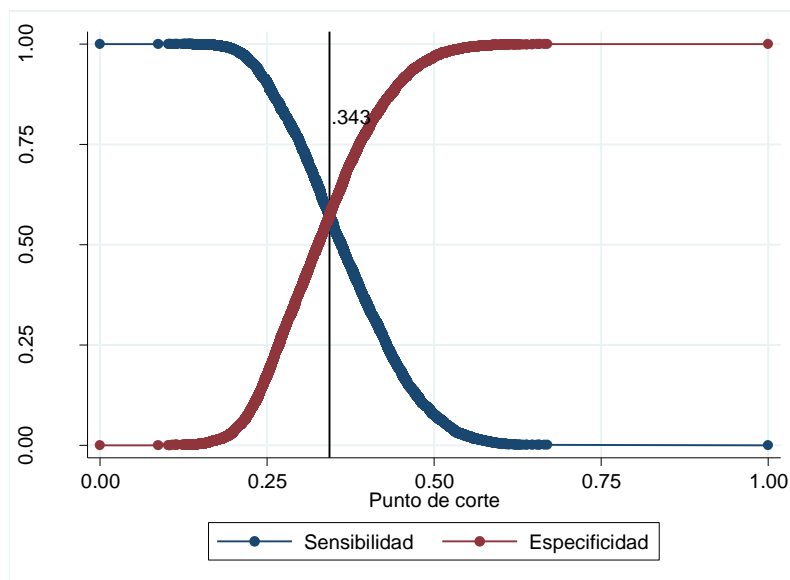
Tabla No. 9
Matriz de Confusión de demanda de salud, punto de corte $\Pr(Y=1)$

Clasificación con $\Pr(D) \geq .3392$	
Sensibilidad	59.10%
Especificidad	55.35%
Falso positivo	44.65%
Falso negativo	40.90%
Porcentaje de positivos incorrectamente clasificados	59.01%
Porcentaje de negativos incorrectamente clasificados	27.94%
Clasificación global	56.64%

Fuente: ENSANUT 2011-2013
Elaboración: Diego Villacreses

Siguiendo con la Ilustración No. 4 se puede observar la sensibilidad y especificidad en relación a todos los puntos de corte entre 0 y 1 para la demanda de salud, se encuentra que la sensibilidad y especificidad llegan a su máxima aproximación en el punto 0.343, información que sirve de insumo en la siguiente matriz de confusión. En el Anexo 2 la Ilustración No. 8 corresponde a la probabilidad de enfermarse y el punto de corte aproximado es 0.411, mientras que la Ilustración No. 10 corresponde la demanda de servicios de salud y tiene como punto de corte 0.185.

Ilustración No. 4
Sensibilidad y especificidad en relación a puntos de corte, demanda de salud



Fuente: ENSANUT 2011-2013

Elaboración: Diego Villacreses

La Tabla No. 10 es la matriz de confusión con el punto de corte donde se cruzan la sensibilidad y la especificidad, como se aprecia la predicción global pasa de 56.64% a 57.18%, la sensibilidad de 59.1% a 57.2% y la especificidad de 55.35% a 57.17%, por lo que se puede considerar que hay una mejora general de la clasificación del modelo. En el Anexo 2 la Tabla No. 19 representa la matriz de confusión con los puntos de corte nuevos para la probabilidad de enfermarse, donde se encuentra que esta tiene una capacidad de predicción global del 64.41%. La Tabla No. 21 es la matriz de confusión con el nuevo punto de corte con una predicción global de 59.79%.

Tabla No. 10
Matriz de Confusión de demanda de salud, punto de corte min(*sens. –espec.*)

Clasificación con $Pr(D) \geq .3433$	
Sensibilidad	57.20%
Especificidad	57.17%
Falso positivo	42.83%
Falso negativo	42.80%
Porcentaje de positivos incorrectamente clasificados	58.80%
Porcentaje de negativos incorrectamente clasificados	28.21%
Clasificación global	57.18%

Fuente: ENSANUT 2011-2013

Elaboración: Diego Villacreses

Con el fin de conocer el efecto que tienen sobre la salud otras variables comportamentales se presenta la Tabla No. 11, donde se estudian las mismas variables que en el modelo anterior pero se añade la actividad física de la persona, esta información se obtiene del Formulario de Actividad Física, cabe recalcar que la información de este formulario únicamente se levantó en el área urbana y para personas entre 18 y 59 años.

Para calcular sí la cantidad de actividad física que realiza una persona es la adecuada se utilizan las recomendaciones de la Organización Mundial de la Salud (2010), donde para el grupo de edad correspondiente a adultos (18 a 64 años) se debe cumplir con:

- Como mínimo 150 minutos semanales a actividad física moderada, o 75 minutos de actividad física vigorosa por semana, o una combinación equivalente de estas actividades.
- La actividad física se deberá practicar en sesiones de 10 minutos de duración, como mínimo.
- Con el fin de obtener mayores beneficios para la salud, se recomienda actividad moderada de 300 minutos por semana, o hasta 150 minutos semanales de actividad física intensa, o una combinación equivalente.
- Dos veces o más por semana, realicen actividades de fortalecimiento de los grandes grupos musculares.

Las tres primeras condiciones se pueden estimar a partir de la ENSANUT 2011-2013, la cuarta no fue tomada en cuenta en el Formulario de Actividad Física, para todos los cálculos se considera únicamente las sesiones de más de 10 minutos. Para los fines del presente estudio se generan 3 categorías, la primera se refiere a las personas que no cumplen con los mínimos requisitos de actividad física recomendados (presentan menos de 150 minutos de actividad física moderada, 75 minutos de actividad vigorosa o una combinación de ambas), las personas que cumplen los requisitos mínimos (entre 150 y menos de 300 minutos de actividad moderada y sus equivalentes) y las que cumplen con los requisitos ideales (300 o más minutos de actividad moderada y sus equivalentes).

Se presenta la Tabla No. 11, donde se consideran las mismas variables independientes que en los modelos anteriores, y se añade la variable categórica de actividad física. Tomando en cuenta que el Formulario de Actividad Física fue realizado solamente en el área urbana, no se incluye la variable de área. Al analizar los resultados se encuentra que a diferencia del modelo anterior la edad, escolaridad del jefe de hogar y el Índice Proxy de Riqueza dejan de ser estadísticamente significativos, es de importancia tomar en cuenta que los coeficientes se mantienen sumamente parecidos al modelo anterior, pero los errores estándares crecen considerablemente, esto se debe a que si bien se tienen 12,186 observaciones, al analizar errores clusterizados no se considera dicho número para el cálculo, sino un ajuste por el total de clústeres, que en este caso son 1031 y en las regresiones anteriores fueron 1700. Nótese que al igual que en la Tabla anterior se encuentra que la presencia de servicios higiénicos en el hogar y un cuarto exclusivo para cocinar son significativos y con signo negativo en los modelos de demanda de salud y probabilidad de enfermarse.

La variable de interés en el análisis actual que es la actividad física es estadísticamente significativa para los modelos sobre la probabilidad de enfermarse y demandar salud, como se puede apreciar la variable es significativa al presentar un cambio de la primera categoría (no cumplir con los requisitos establecidos por la OMS) a la tercera categoría (cumplir con los requisitos ideales). Donde el efecto marginal es -0.0543 para la probabilidad de enfermarse, lo que se interpreta como una disminución de 5.43 puntos porcentuales en la variable de análisis. El cumplimiento de las recomendaciones ideales de actividad física además representan una disminución de 4.02 puntos porcentuales en la probabilidad de demandar salud. Como puede apreciarse la presencia de servicios higiénicos y de un cuarto exclusivo para cocinar acumulan un efecto de 11.5 puntos porcentuales, pero el coste económico de implementar dichos servicios (si bien necesarios) es sumamente alto a nivel nacional, la concientización de la importancia de la actividad física y su impacto en la salud, es relativamente menos costosa.

Tabla No. 11

Regresión Logística de probabilidad enfermarse, demandar salud y demandar servicios de salud. Donde se incluye variable cumplimiento de recomendaciones de actividad física

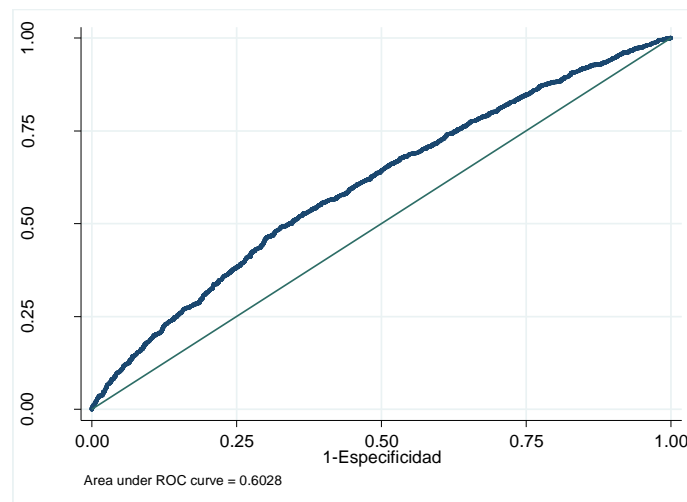
Variables	(1)		(2)		(3)	
	Logit	dy/dx	Logit	dy/dx	Logit	dy/dx
Edad	0.0216 (0.0232)	0.00511 (0.00549)	0.0205 (0.0242)	0.00462 (0.00546)	0.0319 (0.0256)	0.00475 (0.00381)
Edad al cuadrado	-2.15e-05 (0.000312)	-5.10e-06 (7.38e-05)	-1.46e-05 (0.000322)	-3.30e-06 (7.28e-05)	-0.000181 (0.000341)	-2.70e-05 (5.07e-05)
Logaritmo de Ingreso Total Hogar	-0.0894* (0.0473)	-0.0212* (0.0112)	-0.00278 (0.0481)	-0.000629 (0.0109)	-0.0195 (0.0660)	-0.00290 (0.00984)
Índice de Proxy de Riqueza	-0.0574 (0.0509)	-0.0136 (0.0120)	-0.0588 (0.0536)	-0.0133 (0.0121)	0.0479 (0.0626)	0.00714 (0.00931)
Escolaridad Jefe de Hogar	0.0113 (0.00814)	0.00268 (0.00193)	0.0100 (0.00818)	0.00227 (0.00185)	0.0166 (0.0104)	0.00248 (0.00155)
Etnia (base = indígena)	0.805*** (0.195)	0.190*** (0.0446)	0.766*** (0.203)	0.174*** (0.0443)	0.418 (0.314)	0.0629 (0.0445)
Afro/negro/mulato	0.645*** (0.209)	0.151*** (0.0487)	0.457** (0.213)	0.100** (0.0466)	0.195 (0.336)	0.0275 (0.0462)
Montubio	0.353** (0.152)	0.0812** (0.0338)	0.328** (0.165)	0.0706** (0.0337)	0.186 (0.279)	0.0261 (0.0370)
Mestizo/otro	0.616** (0.244)	0.144** (0.0572)	0.556** (0.245)	0.123** (0.0545)	0.262 (0.380)	0.0377 (0.0545)
Blanco	-0.302*** (0.0658)	-0.0716*** (0.0155)	-0.300*** (0.0641)	-0.0678*** (0.0143)	-0.524*** (0.0815)	-0.0781*** (0.0118)
Sexo (hombre =1)	-0.0846* (0.0437)	-0.0200* (0.0103)	-0.117*** (0.0447)	-0.0265*** (0.0101)	-0.105* (0.0557)	-0.0156* (0.00831)
Cantidad personas 0-5 años	-0.0300 (0.0795)	-0.00711 (0.0188)	-0.133* (0.0758)	-0.0301* (0.0171)	-0.0492 (0.0910)	-0.00733 (0.0136)
Cantidad personas 65 más	0.0769 (0.112)	0.0182 (0.0265)	0.0378 (0.105)	0.00855 (0.0237)	-0.131 (0.127)	-0.0196 (0.0189)
Agua Potable	-0.263*** (0.0783)	-0.0624*** (0.0185)	-0.223*** (0.0861)	-0.0504*** (0.0194)	-0.143 (0.102)	-0.0213 (0.0151)
Servicios Higiénicos	0.304* (0.177)	0.0721* (0.0418)	0.249 (0.191)	0.0562 (0.0432)	0.291 (0.234)	0.0434 (0.0348)
Piso de Tierra	-0.222** (0.0913)	-0.0526** (0.0216)	-0.183** (0.0924)	-0.0414** (0.0209)	-0.0806 (0.108)	-0.0120 (0.0161)
Cuarto exclusivo para cocinar	0.0582 (0.0645)	0.0138 (0.0153)	0.0527 (0.0688)	0.0119 (0.0156)	0.277*** (0.0873)	0.0412*** (0.0131)
Seguro Social	0.0582 (0.0645)	0.0138 (0.0153)	0.0527 (0.0688)	0.0119 (0.0156)	0.277*** (0.0873)	0.0412*** (0.0131)
Actividad Física (base=no cumple recomendaciones)	-0.0160 (0.0806)	-0.00385 (0.0193)	0.0199 (0.0863)	0.00459 (0.0199)	0.0153 (0.109)	0.00233 (0.0166)
Recomendación mínima	-0.218*** (0.0636)	-0.0518*** (0.0151)	-0.163** (0.0637)	-0.0369** (0.0144)	-0.0823 (0.0856)	-0.0122 (0.0127)
Recomendación ideal	-0.343 (0.509)		-1.083** (0.528)		-2.079*** (0.670)	
Constante						
Observaciones	12,186	12,186	12,186	12,186	12,186	12,186
Errores estándares Clusterizados en paréntesis						
*** p<0.01, ** p<0.05, * p<0.1						

Fuente: ENSANUT 2011-2013

Elaboración: Diego Villacreses

Como se puede apreciar en la Ilustración No. 5 la capacidad promedio de predicción del modelo de demanda de salud al incluir la variable de actividad física es del 60.28%, lo cual se encuentra dentro de los rangos mínimos aceptables para una regresión de variables binarias. Las curvas ROC para las regresiones sobre la probabilidad de enfermarse y la demanda de servicios de salud se encuentran en el Anexo No. 2 y corresponden a las Ilustración No. 11 y a la Ilustración No. 13 respectivamente, donde se aprecian valores de 60.95% y 61.77% para la capacidad promedio de predicción, lo que también se encuentra dentro del rango aceptable para este estimador.

Ilustración No. 5
Curva ROC de regresión logística sobre demanda de salud al incluir actividad física



Fuente: ENSANUT 2011-2013
Elaboración: Diego Villacreses

Siguiendo con la Tabla No. 12 se encuentra que el modelo de demanda de salud tiene una capacidad de predicción del 52.91% al tomar en cuenta el punto de corte equivalente a la prevalencia de $Y=1$, por lo que no es recomendable este punto de corte. En el Anexo No. 2 se aprecia en la Tabla No. 22 y Tabla No. 24 que corresponden a las Matrices de Confusión de la probabilidad de enfermarse y demanda de servicios de salud con valores para la predicción global de 52.92% y 66.58% respectivamente. Por lo que este punto de corte solamente es recomendable para la tercera regresión.

Tabla No. 12
Matriz de Confusión de demanda de salud al incluir actividad física, punto de corte $Pr(Y=1)$

Clasificación con $Pr(D) \geq .3392$	
Sensibilidad	69.68%
Especificidad	43.01%
Falso positivo	56.99%
Falso negativo	30.32%
Porcentaje de positivos incorrectamente clasificados	58.09%
Porcentaje de negativos incorrectamente clasificados	29.37%
Clasificación global	52.91%

Fuente: ENSANUT 2011-2013

Elaboración: Diego Villacreses

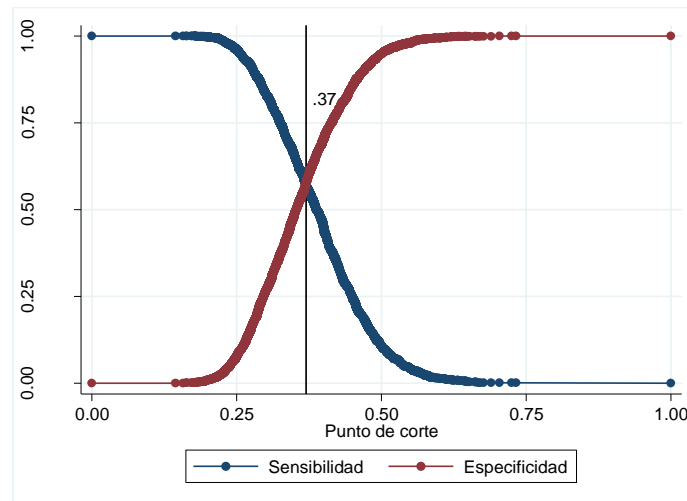
Nótese la Ilustración No. 6, donde se presenta el punto de corte para la predicción de las regresiones logísticas para variables dicotómicas, donde se cruzan las curvas de sensibilidad y especificidad. Para el caso de la demanda de salud el punto de cortes es 0.37, mientras que para la probabilidad de enfermarse es 0.434 (

Ilustración No. **12**) y 0.187 para la demanda de servicios de salud (

Ilustración No. 14).

Ilustración No. 6

Sensibilidad y especificidad en relación a puntos de corte al incluir actividad física, demanda de salud



Fuente: ENSANUT 2011-2013

Elaboración: Diego Villacreses

Al analizar la Matriz de Confusión de demanda de salud (Tabla No. 13), con el punto de corte analizado anteriormente se aprecia que el modelo tiene un punto de corte del 62%. En el Anexo No. 2 se encuentra que la capacidad de predicción de la regresión logística para la probabilidad de enfermarse es de 62% (

Tabla No. 23) y para la demanda de servicios de salud es de 60.24% (Tabla No. 25).

Tabla No. 13

Matriz de Confusión de demanda de salud al incluir actividad física, punto de corte $\min(sens. - espec.)$

Clasificación con $Pr(D) \geq .4343$	
Sensibilidad	29.66%
Especificidad	81.09%
Falso positivo	18.91%
Falso negativo	70.34%
Porcentaje de positivos incorrectamente clasificados	51.94%
Porcentaje de negativos incorrectamente clasificados	33.85%
Clasificación global	62.00%

Fuente: ENSANUT 2011-2013

Elaboración: Diego Villacreses

Modelos de Conteo para análisis de intensidad de enfermedades y demanda de salud

En la sección anterior se analizó la demanda de salud como una variable dicotómica, permitiendo así comprender las variables observables que afectan la probabilidad de demandar salud. Con el fin de extender el análisis se plantea el uso de Modelos de Conteo buscando entender las variables que afectan el tiempo que una persona pasa enferma.

A continuación se presenta un análisis mediante un modelo de conteo en base a una regresión de Poisson, con el fin de conocer los determinantes de la cantidad de días que la persona pasa enfermo, la cantidad de días que pasa enfermo condicionado a que demanda salud y a que demanda servicios de salud específicamente, como se puede apreciar en la Tabla No. 14, los tres análisis corresponden a cada una de las regresiones respectivamente.

Al ser modelos de regresión de Poisson el coeficiente deberá ser analizado como el cambio porcentual generado en la variable dependiente si la variable dependiente aumentase en una unidad si la misma no se encontrase en forma logarítmica, en el caso de ser la variable independiente logarítmica el coeficiente deberá ser interpretado como el aumento porcentual en la Y al aumentar en un 1% la variable independiente. Nótese que en este análisis la educación deja de ser significativa, lo cual es más intuitivo pues ya se encuentran incluidos el ingreso y la riqueza que son variables altamente correlacionadas con la educación.

Del total de variables que anteriormente eran significativas con la nueva especificación, para la cantidad de días enfermo, ahora solo son significativas la edad, el logaritmo del ingreso, el índice proxy de la riqueza, las etnias, el agua, el sexo, el agua potable, si el hogar presenta un cuarto exclusivo para cocinar y el seguro social. Si bien en general se mantienen los mismos signos en la regresión. El ingreso y la riqueza tienen signos negativos, las demás etnias en relación a los indígenas con signo positivo. Nótese que en esta especificación el área ahora tiene signo positivo.

Siguiendo la Tabla No. 14 se encuentra que tanto para la cantidad de días que la persona ha pasado enfermo como la cantidad de días que paso enfermo condicionado a demandar salud y condicionado a demandar servicios de salud la edad tiene un efecto positivo y significativo ya que al aumentar un año de edad la persona aumenta en alrededor de 3 puntos porcentuales la cantidad de días que pasará enfermo. Nótese que el logaritmo del ingreso y el índice proxy de riqueza solamente son significativos en la cantidad de días que pasa enferma la persona, cuando se condiciona a que la persona haya demandado salud deja de ser significativo el índice proxy y cuando se condiciona que haya demandado servicios de salud dejan de ser significativas las dos variables, dicho fenómeno puede ser debido a la presencia de seguro social entre las variables independientes o la alta correlación entre las variables de ingreso, riqueza y estudio.

Como puede apreciarse en la Tabla No. 14 el seguro social es significativo para los tres modelos, igualmente en los tres modelos ser hombre presenta un signo negativo y significativo por lo que se podría afirmar que las mujeres tienden a pasar más días enfermas, no se puede

afirmar que el fenómeno observado se da por una discriminación durante el tiempo en el acceso a servicios de salud a la mujer o a cuestiones biológicas. También se encuentra que las personas autodenominadas indígenas pasan menos tiempo enfermas y demandan menos salud, pero al realizar la regresión para la demanda de servicios de salud las diferencias entre etnias no son estadísticamente significativas excepto en el caso de los montubios. A la vez se encuentra que la presencia de agua potable solamente es significativa al 90% en la regresión de cantidad de días que pasa enfermo y en los restantes modelos deja de ser significativa. Tanto la presencia de servicios higiénicos como el piso de tierra son no significativos los cual es poco intuitivo y podría deberse a la calidad de la información. Finalmente se encuentra que la presencia de un cuarto exclusivo para cocinar solamente es significativo en la cantidad de días que la persona pasa enferma.

Tabla No. 14

Modelo de Poisson de determinantes para días de enfermedad, días de enfermedad condicionado a que la persona demandó salud y condicional a que la persona haya demandado servicios de salud.

Variables	(1)	(2)	(3)
Edad	0.0293*** (0.00598)	0.0312*** (0.00671)	0.0340*** (0.00904)
Edad al cuadrado	-0.000104 (6.71e-05)	-0.000109 (7.45e-05)	-9.34e-05 (9.89e-05)
Logaritmo de Ingreso Total Hogar	-0.0626*** (0.0217)	-0.0548** (0.0240)	-0.0474 (0.0336)
Índice de Proxy de Riqueza	-0.0504* (0.0277)	-0.0519 (0.0328)	-0.000379 (0.0461)
Escolaridad Jefe de Hogar	-7.39e-05 (0.00432)	-0.000138 (0.00479)	0.00124 (0.00673)
Etnia (base = indígena)			
Afro/negro/mulato	0.164* (0.0943)	0.286*** (0.106)	0.102 (0.158)
Montubio	0.333*** (0.0920)	0.439*** (0.106)	0.308** (0.146)
Mestizo/otro	0.163** (0.0738)	0.289*** (0.0853)	0.202 (0.123)
Blanco	0.277** (0.126)	0.422*** (0.144)	0.347 (0.212)
Área (1=Urbano)	0.0770* (0.0432)	0.113** (0.0471)	-0.0247 (0.0622)
Sexo (hombre =1)	-0.337*** (0.0299)	-0.361*** (0.0337)	-0.579*** (0.0439)
Cantidad personas 0-5 años	-0.0223 (0.0224)	-0.0499** (0.0247)	-0.0782** (0.0341)
Cantidad personas 65 más	-0.0284 (0.0378)	-0.0638 (0.0404)	-0.0403 (0.0507)
Agua Potable	0.0873* (0.0494)	0.0749 (0.0532)	0.0981 (0.0730)
Servicios Higiénicos	-0.0175 (0.0434)	-0.0318 (0.0498)	0.0228 (0.0711)
Piso de Tierra	-0.0446 (0.0774)	-0.100 (0.0855)	-0.0758 (0.127)
Cuarto exclusivo para cocinar	-0.101** (0.0469)	-0.0624 (0.0536)	-0.109 (0.0825)
Seguro Social	0.0747** (0.0336)	0.112*** (0.0374)	0.281*** (0.0529)
Constante	0.897*** (0.193)	0.465** (0.217)	-0.00544 (0.300)
Observaciones	30,632	30,632	30,632
Errores estándares Clusterizados en paréntesis			
*** p<0.01, ** p<0.05, * p<0.1			

Fuente: ENSANUT 2011-2013

Elaboración: Diego Villacreses

Al igual que en la sección anterior con el fin de comprender el efecto de otras variables comportamentales se incluirá la actividad física como una variable independiente, para su cálculo se utilizó la misma metodología recomendada por la OMS en el 2010. Se deben tomar

en cuenta las mismas consideraciones, ya que el análisis únicamente se puede dar en el área urbana del país y para personas entre 18 y 59 años.

Como puede apreciarse en la Tabla No. 15 la edad, el Índice Proxy de Riqueza y la escolaridad del jefe de hogar dejan de ser significativos para los tres modelos, y el logaritmo de ingreso total del hogar deja de ser significativo para los dos últimos modelos. Se encuentra que los coeficientes son similares a los presentados en los modelos de la tabla anterior, mientras que los errores estándares para las variables de edad y el logaritmo del ingreso total del hogar son considerablemente más grandes, fenómeno que se puede explicar por el reducido número de observaciones y de clústeres disponibles. También es de interés analizar la variable de seguro social, ya que en la Tabla anterior esta era significativa para los tres modelos mientras que en el presente análisis es únicamente significativo para la demanda de servicios de salud, lo cual se puede deber al reducido número de clústeres. Por otra parte el coeficiente del seguro social es considerablemente más grande en la cantidad de días de enfermedad condicionado a la demanda de servicios de salud (0.213) en relación al tiempo que la persona pasa enferma (0.0428), pese a incluir ingresos y educación en la regresión, se encuentra que el acceso al seguro social aumenta considerablemente el acceso a servicios de salud.

En relación a la variable de actividad física, se confirma el resultado de las regresiones anteriores, pero en este caso se encuentra que al cumplir con las recomendaciones ideales de la OMS se encuentra un efecto significativo en los tres modelos. Se puede observar que cumplir con las recomendaciones ideales de actividad física reduce en promedio en 12.8% la cantidad de días que la persona pasaría enferma asumiendo que el resto de variables se mantienen constantes. En relación a la cantidad de días que la persona pasaría enferma condicionado a demandar salud y demandar servicios de salud se estima que el efecto sería una disminución del 12.7% y el 14.78% respectivamente.

Al igual que en los modelos anteriores se puede apreciar que las variables con mayor impacto sobre la salud para las tres variables independientes son la etnia y el sexo, encontrándose que en general los hombres presentan menor prevalencia y tiempo de enfermedad que las mujeres y que los indígenas reportan la menor probabilidad condicional y tiempo de enfermedad condicionado al conjunto de variables presentado en relación al resto de etnias.

Dentro de las variables que se pueden manipular en base al comportamiento individual o acceso a bienes y servicios, se observa que para la primera variable independiente la presencia de un cuarto exclusivo para cocinar es la variable con el coeficiente más importante (-0.141) seguido de realizar actividad física dentro de las recomendaciones de la OMS (-0.137). Mientras que para las regresiones realizadas para el tiempo de enfermedad condicionado a la demanda de salud y a la demanda de servicios de salud la variables comportamental o de acceso a bienes y servicios más importante es la actividad física.

Es importante aclarar que si bien en el modelo se observa que únicamente la presencia de un cuarto exclusivo para cocinar es estadísticamente significativa, esto se puede deber a la alta correlación entre la presencia de agua potable, servicios higiénicos y piso de tierra en la

vivienda. Si bien las estimaciones afirman que la variable más importante es el cuarto para cocinar, se puede afirmar que todo el conjunto de variables tienen en conjunto un efecto considerablemente grande, que vendría a ser para todas las variables dependientes, el efecto más importante de las variables comportamentales y de acceso a bienes y servicios.

Tabla No. 15

Modelo de Poisson de determinantes para días de enfermedad, días de enfermedad condicionado a que la persona demandó salud y condicional a que la persona haya demandado servicios de salud. Donde se incluye variable cumplimiento de recomendaciones de actividad física

Variables	(1)	(2)	(3)
Edad	0.0176 (0.0182)	0.0166 (0.0201)	0.0333 (0.0282)
Edad al cuadrado	3.49e-05 (0.000234)	6.64e-05 (0.000256)	-0.000111 (0.000361)
Logaritmo de Ingreso Total Hogar	-0.0854** (0.0403)	-0.0461 (0.0438)	-0.0625 (0.0688)
Índice de Proxy de Riqueza	-0.0305 (0.0444)	-0.0395 (0.0508)	0.00767 (0.0714)
Escolaridad Jefe de Hogar	0.00257 (0.00666)	0.00463 (0.00743)	0.0148 (0.0104)
Etnia (base = indígena)			
Afro/negro/mulato	0.262 (0.170)	0.405* (0.210)	0.202 (0.350)
Montubio	0.400** (0.183)	0.483** (0.228)	0.244 (0.374)
Mestizo/otro	0.245* (0.147)	0.357* (0.186)	0.165 (0.320)
Blanco	0.394* (0.205)	0.539** (0.243)	0.330 (0.406)
Sexo (hombre =1)	-0.218*** (0.0504)	-0.244*** (0.0567)	-0.459*** (0.0826)
Cantidad personas 0-5 años	-0.00681 (0.0387)	-0.0573 (0.0455)	-0.0898 (0.0591)
Cantidad personas 65 más	0.0463 (0.0596)	-0.0577 (0.0650)	-0.00265 (0.0912)
Agua Potable	0.0375 (0.0981)	0.0280 (0.110)	0.106 (0.144)
Servicios Higiénicos	-0.0317 (0.0633)	-0.0563 (0.0748)	-0.00711 (0.109)
Piso de Tierra	0.0211 (0.151)	-0.0560 (0.167)	0.0563 (0.258)
Cuarto exclusivo para cocinar	-0.141** (0.0665)	-0.131* (0.0768)	-0.154 (0.124)
Seguro Social	0.0428 (0.0532)	0.0520 (0.0623)	0.213** (0.0934)
Actividad Física (base=no cumple recomendaciones)			
Recomendación mínima	-0.0627 (0.0622)	-0.0452 (0.0728)	-0.0266 (0.106)
Recomendación ideal	-0.137*** (0.0518)	-0.136** (0.0588)	-0.160* (0.0837)
Constante	1.335*** (0.463)	0.878* (0.521)	0.116 (0.814)
Observaciones	12,186	12,186	12,186
Errores estándares Clusterizados en paréntesis			
*** p<0.01, ** p<0.05, * p<0.1			

Fuente: ENSANUT 2011-2013

Elaboración: Diego Villacreses

Modelos de Conteo para análisis de intensidad de enfermedades y demanda de salud con un enfoque territorial

Los análisis realizados en la presente disertación no profundizan en el estudio del efecto de variables territoriales a parte del efecto de vivir en el área urbana en relación al área rural. Con el fin de conocer con mayor detalle los diferentes resultados de vivir en la región Costa en relación a la región Sierra del Ecuador, se presentan dos Tablas, la primera compara tres regresiones con las diferentes variables dependientes utilizadas en los modelos y Tablas anteriores en la Sierra y en la Costa como modelos separados, donde el conjunto de variables independientes será el mismo utilizado en los modelos anteriores (sin actividad física). La segunda Tabla será igual a la anterior exceptuando que en este caso se incluirá la variable de actividad física, utilizando la misma metodología que anteriormente y se deberán tomar las mismas consideraciones que en los modelos anteriores (se dispone información para el área urbana y población entre 18 y 59 años).

Se presentan las tres Tablas en base a regresiones de Poisson o modelos de conteo, ya que estas explican de forma más amplia la información de demanda de salud, considerándose la forma funcional y las recomendaciones de Grossman (1972). Si bien es de interés conocer a un mayor grado de profundidad los efectos territoriales, se podría plantear la comparación de regresiones entre la Quito y Guayaquil que se encuentran autorepresentadas en la ENSANUT 2011-2013, pero hay que tomar en cuenta que estas desagregaciones territoriales son las mínimas aceptadas por el diseño muestral por lo que un análisis de determinantes que puede ser comprendido como una desagregación de estimadores dentro del territorio analizado no sería representativo, y sus coeficientes no serían interpretables. Razones por las que en la que en la Tabla No. 16 se considera como variable de control la Región Natural y como se puede apreciar la diferencia es significativa entre Sierra y Costa para la cantidad de días que la persona pasa enferma y la misma característica controlada por la demanda de servicios de salud, aunque no es significativa para el tiempo de enfermedad condicionado a demandar salud.

Siguiendo con la Tabla No. 16 se encuentra que las regiones del Oriente e Insular no presentan una diferencia significativa con la Sierra en las 3 regresiones. Como se esperaba los resultados en el resto de variables se mantienen considerablemente parecidos a las regresiones presentadas en la Tabla No. 14 confirmándose la robustez de los resultados. Por ejemplo se encuentra que las variables de logaritmo del ingreso total del hogar y edad se mantienen iguales hasta al tercer decimal para las tres variables dependientes. Además variables importantes como el área, sexo, agua potable mantienen su significancia, el mismo signo y una intensidad similar. A partir del análisis presentado se puede concluir la robustez de los resultados en relación al ingreso de variables territoriales, y además una diferencia significativa entre la Sierra y la Costa, encontrándose que en la Costa presenta un promedio de días de enfermedad mayor que en la Sierra, razón por la que se presenta la Tabla No. 17, donde se presentan las 3 regresiones que se han analizado a lo largo de la presente disertación utilizando un modelo de Poisson para variables de conteo, donde las variables dependientes

serán la cantidad de días que la persona paso enferma. En este caso se comparan las estimaciones las regresiones entre la región Sierra y Costa.

Tabla No. 16

Modelo de Poisson de determinantes para días de enfermedad, días de enfermedad condicionado a que la persona demandó salud y condicional a que la persona haya demandado servicios de salud. Se añade variable de control territorial: Región Natural.

Variables	(1)	(2)	(3)
Edad	0.029 (0.006)***	0.031 (0.007)***	0.034 (0.009)***
Edad al cuadrado	-0.000 (0.000)	-0.000 (0.000)	-0.000 (0.000)
Logaritmo de Ingreso Total Hogar	-0.062 (0.022)***	-0.054 (0.024)**	-0.047 (0.034)
Índice de Proxy de Riqueza	-0.049 (0.028)*	-0.052 (0.033)	-0.001 (0.046)
Escolaridad Jefe de Hogar	-0.001 (0.004)	-0.000 (0.005)	0.002 (0.007)
Etnia (base = indígena)			
Afro/negro/mulato	0.124 (0.097)	0.253 (0.108)**	0.160 (0.160)
Montubio	0.282 (0.098)***	0.396 (0.111)***	0.386 (0.154)**
Mestizo/otro	0.138 (0.075)*	0.266 (0.086)***	0.240 (0.123)*
Blanco	0.241 (0.130)*	0.392 (0.146)***	0.400 (0.214)*
Área (1=Urbano)	0.049 (0.046)	0.091 (0.050)*	0.017 (0.063)
Sexo (hombre =1)	-0.340 (0.030)***	-0.364 (0.034)***	-0.576 (0.044)***
Cantidad personas 0-5 años	-0.021 (0.022)	-0.048 (0.025)**	-0.081 (0.034)**
Cantidad personas 65 más	-0.032 (0.038)	-0.067 (0.040)*	-0.035 (0.050)
Agua Potable	0.095 (0.050)*	0.079 (0.054)	0.091 (0.073)
Servicios Higiénicos	0.007 (0.047)	-0.012 (0.054)	-0.014 (0.075)
Piso de Tierra	-0.038 (0.078)	-0.099 (0.086)	-0.079 (0.128)
Cuarto exclusivo para cocinar	-0.090 (0.047)*	-0.054 (0.054)	-0.124 (0.084)
Seguro Social	0.076 (0.034)**	0.113 (0.037)***	0.279 (0.053)***
Región (1=Sierra)			
Costa	0.079 (0.043)*	0.057 (0.047)	-0.105 (0.064)*
Oriente	0.015 (0.041)	-0.036 (0.044)	0.051 (0.056)
Insular	0.025 (0.071)	-0.030 (0.081)	-0.185 (0.103)*
Constante	0.885 (0.193)***	0.460 (0.217)**	-0.001 (0.301)
Observaciones	30,632	30,632	30,632
Errores estándares Clusterizados en paréntesis			
*** p<0.01, ** p<0.05, * p<0.1			

Fuente: ENSANUT 2011-2013

Elaboración: Diego Villacreses

Como se puede observar en la Tabla No. 17 se presentan dos secciones de regresiones, la primera corresponde a los resultados para las tres variables dependientes usualmente analizadas en la región Sierra, mientras que la segunda presenta las mismas regresiones pero en este caso con resultados para la región Costa. Como se puede observar la variable de edad se mantiene significativa para tanto en la Sierra como en la Costa, pero el efecto de la edad es más fuerte en la primera. Se encuentra que la variable de edad es significativa para la cantidad de días que la persona pasa enferma tanto al ser variable incondicionada, como al ser condicionada a demandar salud y a demandar servicios de salud, para los tres casos se encuentra que el efecto de la edad es más intenso en la Sierra que en la Costa, cabe recalcar que se controla al modelo por área urbano-rural. Como se puede observar en la Tabla No. 17 la mayoría de variables de las dejan de ser significativas al tomar modelos individuales para cada región de interés, lo cual puede deberse a la poca cantidad de observaciones y clústeres para el cálculo de los errores estándares, lo cual infla considerablemente a los mismos.

Las variables significativas en las dos regiones son edad, sexo y seguro social; condiciones que limitan el análisis territorial a las conclusiones obtenidas en la Tabla No. 16 y la información que se pueda obtener de las tres variables significativas. Se encuentra que para la cantidad de días que la persona pasa enfermo la diferencia en el efecto generado por el género tiene una pequeña diferencia entre la Sierra y la Costa (0.002), la diferencia en los efectos entre la cantidad de días de enfermedad condicionado a demandar salud es igualmente pequeño (0.032). Donde se puede evidenciar un efecto considerable es en la variable dependiente condicionada a demandar servicios de salud, ya que la diferencia es equivalente a 0.121, encontrándose que en la Costa las mujeres tienen mayor dificultad para acceder a servicios de salud que en la Sierra, cabe recalcar que dichos coeficientes se encuentran controlados por educación e ingresos.

Finalmente se puede analizar el efecto de la variable de Seguro Social para el tiempo de enfermedad condicionado a demandar servicios de salud, y puede apreciarse que el coeficiente de la misma variable es alrededor del doble en la Costa (0.368) en relación a la Sierra (0.189), lo que puede deberse a una mayor dificultad de acceso a servicios de salud en dicha Región Natural, cabe recalcar que la regresión se encuentra controlada por condiciones de riqueza, ingresos y escolaridad, por lo que se puede considerar al resultado como un coeficiente insesgado. Este análisis no puede ser tomado en cuenta para las dos primeras variables dependientes ya que la variable de Seguro Social no es significativa en ambas regiones para dichos modelos.

Tabla No. 17

Modelo de Poisson de determinantes para días de enfermedad, días de enfermedad condicionado a que la persona demandó salud y condicional a que la persona haya demandado servicios de salud. Se consideran modelos para la Costa (1) y para la Sierra (2).

Variables	(1)			(2)		
	(1)	(2)	(3)	(1)	(2)	(3)
Edad	0.042 (0.008)***	0.045 (0.010)***	0.039 (0.013)***	0.017 (0.009)**	0.019 (0.010)*	0.027 (0.014)**
Edad al cuadrado	-0.000 (0.000)**	-0.000 (0.000)**	-0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	-0.000 (0.000)
Logaritmo de Ingreso Total Hogar	-0.044 (0.033)	-0.028 (0.035)	-0.005 (0.046)	-0.074 (0.031)**	-0.072 (0.035)**	-0.083 (0.053)
Índice de Proxy de Riqueza	-0.075 (0.035)**	-0.079 (0.039)**	-0.029 (0.050)	-0.021 (0.046)	-0.021 (0.057)	0.039 (0.088)
Escolaridad Jefe de Hogar	0.007 (0.006)	0.004 (0.006)	0.006 (0.008)	-0.008 (0.006)	-0.005 (0.007)	-0.005 (0.011)
Etnia (base = indígena)						
Afro/negro/mulato	0.119 (0.181)	0.050 (0.178)	0.120 (0.246)	-0.129 (0.268)	0.215 (0.325)	-0.341 (0.393)
Montubio	0.515 (0.301)*	0.668 (0.340)**	0.754 (0.413)*	0.024 (0.270)	0.284 (0.323)	-0.197 (0.387)
Mestizo/otro	0.189 (0.094)**	0.341 (0.110)***	0.390 (0.165)**	-0.116 (0.261)	0.155 (0.316)	-0.331 (0.375)
Blanco	0.394 (0.203)*	0.531 (0.236)**	0.559 (0.351)	-0.068 (0.292)	0.254 (0.342)	-0.171 (0.424)
Área (1=Urbano)	0.023 (0.063)	0.059 (0.068)	0.011 (0.085)	0.063 (0.075)	0.127 (0.082)	0.071 (0.103)
Sexo (hombre =1)	-0.345 (0.039)***	-0.375 (0.045)***	-0.522 (0.059)***	-0.343 (0.046)***	-0.354 (0.052)***	-0.643 (0.070)***
Cantidad personas 0-5 años	-0.088 (0.034)**	-0.112 (0.039)***	-0.118 (0.046)**	0.033 (0.031)	0.003 (0.033)	-0.050 (0.054)
Cantidad personas 65 más	-0.022 (0.056)	-0.073 (0.053)	-0.113 (0.069)	-0.041 (0.053)	-0.064 (0.059)	0.026 (0.074)
Agua Potable	0.097 (0.082)	0.104 (0.089)	0.236 (0.110)**	0.086 (0.075)	0.047 (0.079)	-0.018 (0.114)
Servicios Higiénicos	-0.009 (0.073)	-0.013 (0.082)	-0.077 (0.100)	0.013 (0.062)	-0.019 (0.072)	0.001 (0.110)
Piso de Tierra	-0.066 (0.100)	-0.131 (0.110)	-0.079 (0.148)	0.005 (0.123)	-0.045 (0.134)	-0.040 (0.215)
Cuarto exclusivo para cocinar	-0.179 (0.073)**	-0.163 (0.079)**	-0.114 (0.120)	-0.053 (0.065)	-0.008 (0.077)	-0.141 (0.124)
Seguro Social	0.024 (0.049)	0.033 (0.052)	0.189 (0.069)***	0.114 (0.049)**	0.178 (0.056)***	0.368 (0.086)***
Constante	0.571 (0.274)**	0.089 (0.290)	-0.580 (0.407)	1.530 (0.365)***	0.943 (0.435)**	0.915 (0.555)*
Observaciones	14,362	14,362	14,362	8,249	8,249	8,249
Errores estándares Clusterizados en paréntesis						
*** p<0.01, ** p<0.05, * p<0.1						

Fuente: ENSANUT 2011-2013

Elaboración: Diego Villacreses

Conclusiones

Según la ENSANUT 2011-2013 el 42.8% de la población ecuatoriana tuvo algún problema de salud el último mes, dicha proporción se encuentra claramente diferenciada entre género pues el 39.7% de los hombres tuvieron un problema de salud frente al 45.7% de las mujeres. Dicha diferencia se puede ver explicada mayoritariamente por las enfermedades de la mujer y el hombre, al eliminar dichas enfermedades del análisis se encuentra que el 41.3% de las mujeres se enfermaron el último mes en relación al 39.1% de los hombres, aunque la diferencia sigue siendo estadísticamente significativa. Según Grossman (1972) dicha diferencia se puede deber a una discriminación de acceso al Capital de Salud a lo largo de la vida de la mujer o a una posible diferencia intrínseca a la resistencia a enfermedades entre los dos géneros. Se puede concluir que la mujer tiene una mayor probabilidad de enfermarse y un tiempo promedio de enfermedad mayor que el del hombre asumiendo iguales condiciones socio-económicas y de actividad física.

Tras realizar los modelos de regresión como es sugerido por Grossman se encuentra el mismo error empírico el autor, la educación tiene el signo contrario al esperado en base al modelo teórico, según el cual las personas más educadas son productores de capital de salud más eficientes, por lo que el signo esperado de la misma debe ser positivo en relación a la cantidad de días saludables que pasa la persona. Dicha contradicción se debe a la especificación del modelo por usar información de corte transversal y no tomar en cuenta el efecto del Capital de Salud del periodo anterior ($t-1$) (Grossman, 1972). Dicha hipótesis es parcialmente confirmada por Wagstaff (2002), quien encuentra que al utilizar modelos de Variables Latentes Recursivos se corrige dicha contradicción entre el modelo teórico y el empírico. Modelo que no puede ser estimado a partir de la información presente en la ENSANUT, ya que esta encuesta es un corte transversal.

Además se confirman el resto de las hipótesis de Grossman, encontrándose que la riqueza y los ingresos generan que las personas tengan mejor salud, y como es intuitivo a medida que aumenta la edad disminuye la cantidad de tiempo que la persona pasa saludable. Dichos resultados son robustos pues se encuentran los mismos signos bajo varias especificaciones y con diferentes variables de control.

Cuando se utilizan modelos de conteo se puede afirmar que la educación deja de ser significativa en el modelo; y, en este caso tiene el signo correcto, la variable podría ser significativa con un tamaño de muestra más grande, por lo que el uso de variables de conteo podría ser una alternativa viable para corregir el signo de la variable de educación en una muestra más grande. Se concluye que tanto la riqueza como el ingreso del hogar mejoran la salud de los individuos, pero como se enunció en el Marco Teórico esta es una relación simultánea pues los ingresos afectan a la salud pero a la vez la salud mejoran ingresos.

Al estudiar características de higiene del hogar como la presencia de agua potable, servicios higiénicos, piso de tierra o cuarto exclusivo para cocinar se puede apreciar que para los modelos de variable binaria la presencia de servicios higiénicos en el hogar es una de las

variables con el coeficiente más importante en relación a disminuir la probabilidad de enfermarse o demandar salud, seguida de cerca por la presencia de un cuarto exclusivo para cocinar. Cuando se analiza la intensidad de la demanda, ósea la cantidad de días que la persona pasa enferma, el resultado en relación a estas características de la vivienda deja de ser significativo por lo que no se puede ser altamente concluyente. La variable que en todas las especificaciones resulta ser significativa y disminuir de forma importante tanto la probabilidad de enfermarse como la cantidad de días que la persona pasa enferma es el cuarto exclusivo para cocinar, por lo que se puede tomar dicho resultado como una conclusión para política pública. Cabe recalcar que variables importantes como el acceso a agua potable y servicios higiénicos no son significativas en todos los modelos ya que se encuentran altamente correlacionadas con el resto de variables del modelo por lo que no deben ser ignoradas al momento de la toma de decisiones.

En relación a la variable de actividad física, se encuentra que el cumplimiento de las recomendaciones ideales (300 minutos a la semana de ejercicio moderada o 150 de ejercicio vigorosa) de actividad tienen un efecto considerable para la salud bajo todas las especificaciones de los modelos, por lo que la ejercitación para adultos entre 18 y 59 años de edad es importante para disminuir la prevalencia de enfermedades. Finalmente se puede concluir que existen diferencias territoriales entre la Costa y la Sierra, tanto en prevalencia de enfermedad, demanda de salud y demanda de servicios de salud, como el tiempo de enfermedad. Encontrándose que el efecto de la edad, el género y el acceso a Seguro Social es más marcado en la Costa que en la Sierra.

Recomendaciones

Se recomienda tener una fuente de información de calidad en relación a aspectos de salud con el fin de tomar decisiones de política acertadas, según la literatura la mejor fuente de información en este caso sería un panel de observaciones, por lo que se recomienda el levantamiento del mismo, con el fin de realizar la especificación sugerida por Wagstaff (2002) y conocer de forma más precisa los coeficientes de las variables analizadas sobre la demanda de salud.

Tras encontrar resultados robustos bajo diferentes especificaciones de los modelos econométricos con diferentes variables de control se puede recomendar como política pública el mejoramiento del equipamiento de la vivienda con especial énfasis en presencia de un cuarto exclusivo para cocinar. Además se recomienda la generación de campañas de concientización en relación a la importancia de la actividad física para la salud, siendo esta la variable de fácil mejoramiento más importante hallada en la presente disertación.

Bibliografía

- Alsan, M., Bloom, D. E., Canning, D., & Jamison, D. (2007). The Consequences of Population Health for Economic Performance. *Health Economic Development, and Household Poverty*, 21-39.
- Andersen, R., & Anderson, O. W. (1968). A Decade of Health Services: Social Survey Trends in Use and Expendi. *University of Chicago Press*.
- Angrist, A., & Pischke, J. (2009). *Mostly Harmless Econometrics*. Princeton, Nueva Jersey: Princeton University Press.
- Berman, P., Kendall, C., & Bhattacharyya, K. (1994). The household production of health: Integrating social science perspectives on micro-level health determinants. *Social Science & Medicine*.
- Bloom, D. E., & Canning, D. (2009). Population Health and Economic Growth. In M. Spence, & M. Lewis, *Health and Growth* (pp. 53-75). Washington: The World Bank.
- Cameron, A. C., & Trivedi, P. K. (1998). *Regression analysis of count data* (1ra ed.). Cambridge University Press.
- Cameron, A., & Trivedi, P. (2009). *Microeconometrics using STATA*. College Station: Stata Press.
- Culyer, A., & Newhouse, J. (2000). *Handbook of health economics. Volume 1A*. Amsterdam: Elsevier Science B.V.
- Elkan, C. (2010). *Predictive analytics and data mining*. Retrieved 2015 Diciembre, from <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.164.416>
- Farmer, P. (1999). Infections and Inequalities: The modern plagues. *Berkeley: University of California Press*.
- Fehrmann, L., Lehtonen, A., Kleinn, C., & Tomppo, E. (2008). Comparison of linear and mixed-effect regression models and a k-nearest neighbour approach for estimation of single-tree biomass. *NRC Research Press*.
- Freire, W. B., Ramírez, M., Belmont, P., Mendieta, M. J., Silva, M. K., Romero, N., et al. (2013). *RESUMEN EJECUTIVO. TOMO I, Encuesta Nacional de Salud y Nutrición del Ecuador. ENSANUT-ECU 2011-2013*. Quito, Ecuador: Ministerio de Salud Pública/Instituto Nacional de Estadística y Censos.
- Greene, W. (2010). *Econometric Analysis*. Prentice Hall.
- Grossman, M. (1972). The Demand for Health: A Theoretical and Empirical Investigation. *National Bureau of Economic Research*.

- Hamoudi , A., & Sachs, J. D. (1999). Economic Consequences of Health Status: A Review of the Evidence. *Center for Internacional Development at Harvard University*.
- Instituto Nacional de Estadísticas y Censos. (2013a). *Actividades y Recursos de Salud*. Retrieved Diciembre 15, 2014, from <http://www.ecuadorencifras.gob.ec/actividades-y-recursos-de-salud/>
- Instituto Nacional de Estadísticas y Censos. (2013b). *Estadísticas de Camas y Egresos Hospitalarios*. Retrieved Diciembre 15, 2014, from <http://www.ecuadorencifras.gob.ec/camas-y-egresos-hospitalarios/>
- Instituto Nacional de Estadísticas y Censos. (2014b, 09 4). *Salud, Salud Reproductiva y Nutrición*. Retrieved 05 06, 2015, from <http://www.ecuadorencifras.gob.ec/salud-salud-reproductiva-y-nutricion/>
- Interuniversity Consortium for Political and Social Research. (n.d.). *Survey of Health Services Utilization and Expenditures, 1963 (ICPSR 7741)*. Retrieved Mayo 04, 2015, from <https://www.icpsr.umich.edu/icpsrweb/NACDA/studies/7741?paging.startRow=276>
- Johnson , R., & Wichern , D. (2007). *Applied Multivariate Statistical Analysis*. Pearson Education, Inc.
- Jones, A. (2000). *Handbook of Health Economics, Volume 1, Edited by A.J. Culyer and J.P. Newhouse*. Elsevier Science B.V.
- Jones, A. M. (2005). *Applied Econometrics for Health Economist. A Practial Guide*. Department of Economics and Related Studies.
- Jones, A. M., Rice, N., Bago, T., & Balia, S. (2013). *Applied health economics*. New York, USA: Routledge.
- Karazsia, B. T., & van Dulmen, H. M. (2008). *Regression Models for Count Data: Illustrations using Longitudinal Predictors of Childhood Injury*.
- Kendall , C. (1990). Public health and the domestic domain: lessons from anthropological research on Diarrheal diseases. *Anthropology and Primary Health Care*.
- Kingman, J. (1993). *Poisson Processes*. Oxford: Oxford University Press.
- Lohr, S. (2009). *Sampling: Desing and Analysis*. Advanced Series.
- McConnell , C., Brue, S., & Flynn, S. (2012). *Economics: Principles, Problems, and Policies*. The McGraw-Hill Series in Economics.
- Muurinen, J. M. (1982). *An economic model of health behaviour : with empirical applications to Finnish health survey data*. University of York.
- Organización Mundial de la Salud. (2010). *Recomendaciones mundiales sobre actividad física para la salud*. Zurich, Suiza.

- Rencher, A. (2002). *Methods of Multivariate Analysis*. Wiley Interscience.
- Samuelson, P. (1954). The Pure Theory of Public Expenditure. *The Review of Economics and Statics*, Vol. 36, No. 4., 387-389.
- STATA CORP. (2014). *Stata user's guide*. College Station, TX: Stata Press.
- Stiglitz, J. (2000). *Economía del sector público*. Antoni Bosch Editor.
- Wagstaff, A. (1986). The demand for health: some new empirical evidence. *Journal of Health Economics*, 195-233.
- Wagstaff, A. (2002). *The demand for health: an empirical reformulation of the Grossman model*. Hoboken, USA: Wiley Series.
- Wagstaff, A. (2002). The Demand for Health: an Empirical Reformulation of the Grossman Model. In A. Jones, & O. O'Donnell, *Econometric Analysis of Health Data* (pp. 15-23). Baffins Lane, Chichester,: John Wiley & Sons.
- Wasserman, L. (2004). *All of Statics. A Concise Course in Statical Inference*. Springer Texts in Statics.
- Wasserman, L. (2004). *All of Statics. A Concise Course in Statics*. Springer Texts in Statistics.
- Wooldridge, J. M. (2002). *Econometric Analysis of Cross Section and Panel Data*. The MIT Press.
- Wooldridge, J. M. (2010). *Introduction to Econometrics*. Cengage Learning Editores.

Anexos

Anexo No. 1

Demostración mediante simulaciones Monte Carlo de igualdad de coeficientes de regresiones con variables dependientes con y sin errores de medida (Código de programación en STATA)

```
clear all
forval y = 1 /100{
clear
set obs 5000
    *se escogen 5000 observaciones para cumplir con propiedades de muestra grande

**Proceso generador de datos
forval x = 1/5{
gen x`x' = rnormal()
}
gen y = 5 + x1*5+x2*3+x3*4+x4*4+x5*9+rnormal()

*Regresión con variable sin error de medida
reg y x*

gen y_a = y +rnormal()
    *Se genera variable con error de medida independiente de Xi

*Regresión con variable con error de medida
reg y_a x*
}
*Se encuentra que en las 100 simulaciones los coeficientes entre la regresion con variable dependiente sin error de medida y la
regresion con variable dependiente con error de medida los coeficientes beta son iguales
```

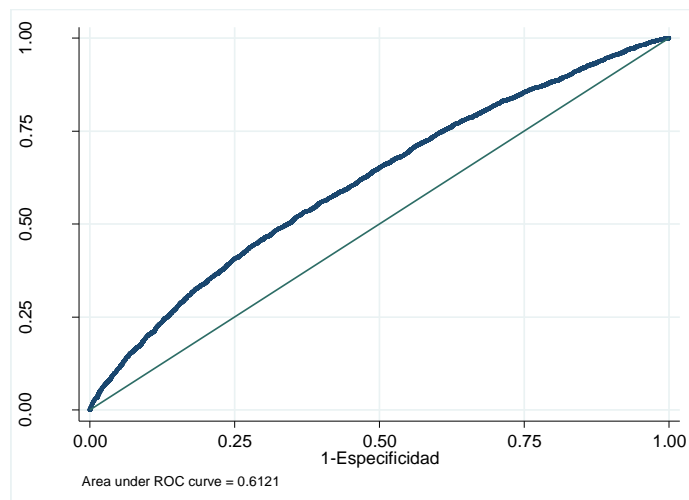
Elaboración: Diego Villacreses

Anexo No. 2

Matrices de confusión, curva de sensibilidad y curva ROC para modelo de probabilidad de enfermarse.

Ilustración No. 7

Curva ROC de regresión logística sobre probabilidad de enfermarse



Fuente: ENSANUT 2011-2013

Elaboración: Diego Villacreses

Tabla No. 18

Matriz de Confusión de probabilidad de enfermarse, punto de corte $\text{Pr}(Y=1)$

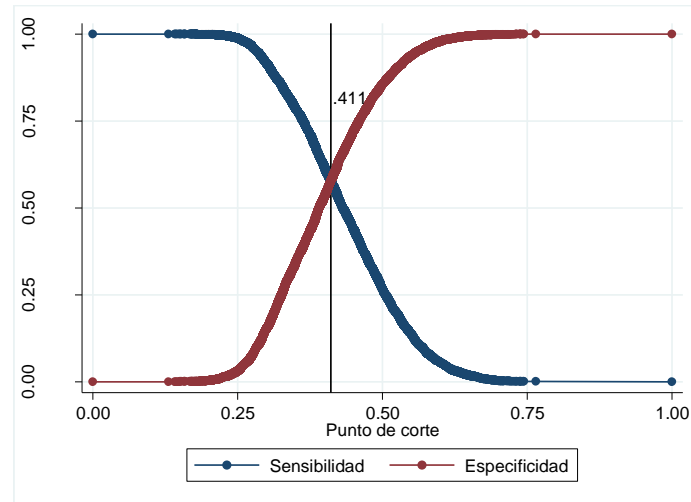
Clasificación con $\text{Pr}(D) \geq .3392$	
Sensibilidad	59.10%
Especificidad	55.35%
Falso positivo	44.65%
Falso negativo	40.90%
Porcentaje de positivos incorrectamente clasificados	59.01%
Porcentaje de negativos incorrectamente clasificados	27.94%
Clasificación global	56.64%

Fuente: ENSANUT 2011-2013

Elaboración: Diego Villacreses

Ilustración No. 8

Sensibilidad y especificidad en relación a puntos de corte, probabilidad de enfermarse



Fuente: ENSANUT 2011-2013

Elaboración: Diego Villacreses

Tabla No. 19

Matriz de Confusión de probabilidad de enfermarse, punto de corte $\min(sens. - espec.)$

Clasificación con $Pr(D) \geq .4112$	
Sensibilidad	31.44%
Especificidad	81.71%
Falso positivo	18.29%
Falso negativo	68.56%
Porcentaje de positivos incorrectamente clasificados	52.57%
Porcentaje de negativos incorrectamente clasificados	30.57%
<u>Clasificación global</u>	<u>64.41%</u>

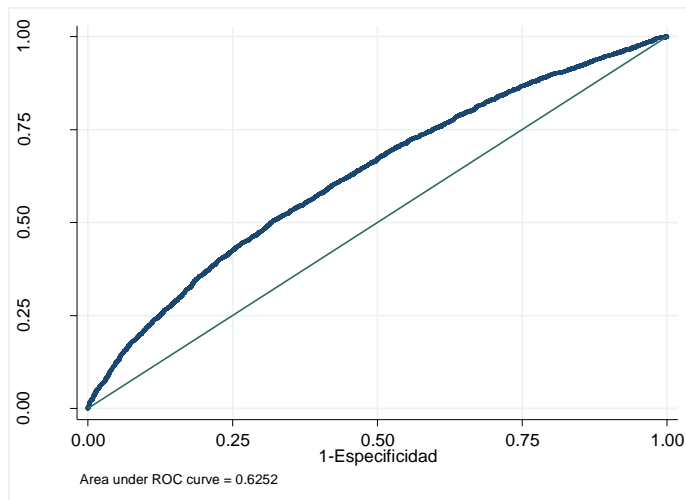
Fuente: ENSANUT 2011-2013

Elaboración: Diego Villacreses

Matrices de confusión, curva de sensibilidad y curva ROC para modelo de demanda de servicios de salud.

Ilustración No. 9

Curva ROC de regresión logística sobre demanda de servicios de salud



Fuente: ENSANUT 2011-2013

Elaboración: Diego Villacreses

Tabla No. 20

Matriz de Confusión de demanda de servicios de salud, punto de corte $\Pr(Y=1)$

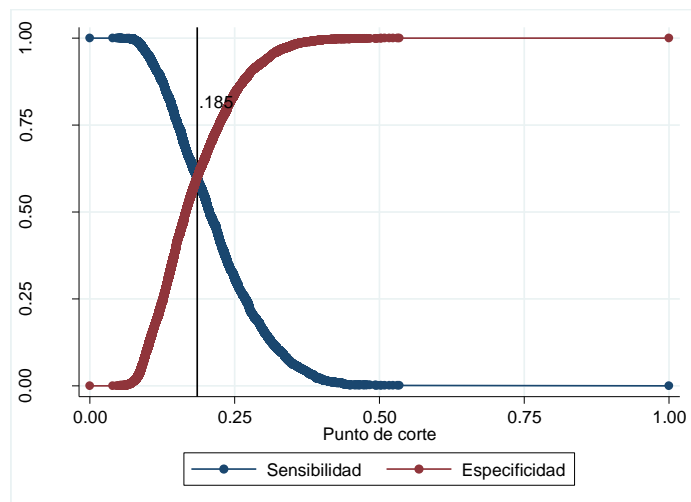
Clasificación con $\Pr(D) \geq .2114$	
Sensibilidad	47.94%
Especificidad	70.79%
Falso positivo	29.21%
Falso negativo	52.06%
Porcentaje de positivos incorrectamente clasificados	72.71%
Porcentaje de negativos incorrectamente clasificados	14.40%
Clasificación global	66.53%

Fuente: ENSANUT 2011-2013

Elaboración: Diego Villacreses

Ilustración No. 10

Sensibilidad y especificidad en relación a puntos de corte, demanda de servicios de salud



Fuente: ENSANUT 2011-2013

Elaboración: Diego Villacreses

Tabla No. 21

Matriz de Confusión demanda de servicios de salud, punto de corte $\min(sens. - espec.)$

Clasificación con $Pr(D) \geq 0.1854$	
Sensibilidad	59.80%
Especificidad	59.78%
Falso positivo	40.22%
Falso negativo	40.20%
Porcentaje de positivos incorrectamente clasificados	74.62%
Porcentaje de negativos incorrectamente clasificados	13.33%
Clasificación global	59.79%

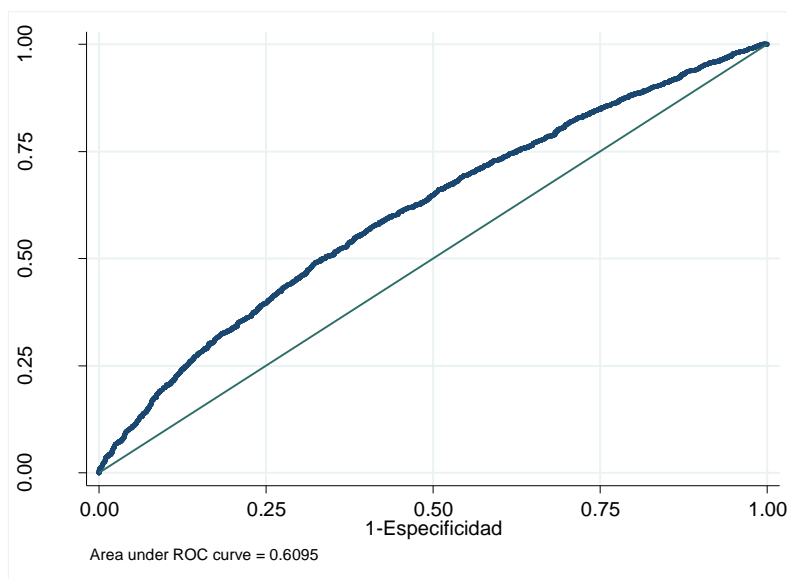
Fuente: ENSANUT 2011-2013

Elaboración: Diego Villacreses

Matrices de confusión, curva de sensibilidad y curva ROC para modelo de probabilidad de enfermarse al incluir variable de actividad física.

Ilustración No. 11

Curva ROC de regresión logística sobre probabilidad de enfermarse al incluir variable de actividad física



Fuente: ENSANUT 2011-2013

Elaboración: Diego Villacreses

Tabla No. 22

Matriz de Confusión de probabilidad de enfermarse al incluir variable de actividad física, punto de corte $\Pr(Y=1)$

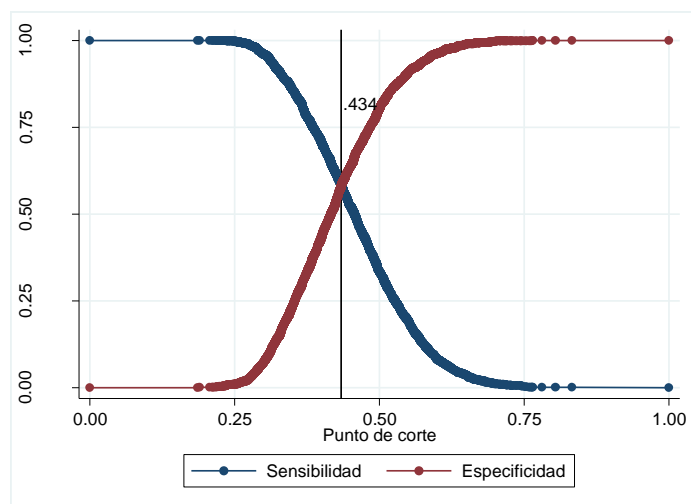
Clasificación con $\Pr(D) \geq .3392$	
Sensibilidad	69.68%
Especificidad	43.01%
Falso positivo	56.99%
Falso negativo	30.32%
Porcentaje de positivos incorrectamente clasificados	58.09%
Porcentaje de negativos incorrectamente clasificados	29.37%
Clasificación global	52.91%

Fuente: ENSANUT 2011-2013

Elaboración: Diego Villacreses

Ilustración No. 12

Sensibilidad y especificidad en relación a puntos de corte, probabilidad de enfermarse, se incluye variable de actividad física



Fuente: ENSANUT 2011-2013
Elaboración: Diego Villacreses

Tabla No. 23

Matriz de Confusión de probabilidad de enfermarse al incluir variable de actividad física, punto de corte $\Pr(Y=1)$

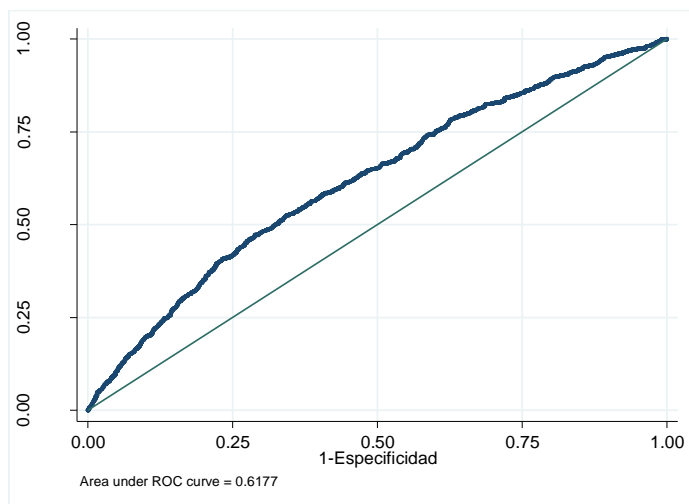
Clasificación con $\Pr(D) \geq .4343$	
Sensibilidad	29.66%
Especificidad	81.09%
Falso positivo	18.91%
Falso negativo	70.34%
Porcentaje de positivos incorrectamente clasificados	51.94%
Porcentaje de negativos incorrectamente clasificados	33.85%
Clasificación global	62.00%

Fuente: ENSANUT 2011-2013
Elaboración: Diego Villacreses

Matrices de confusión, curva de sensibilidad y curva ROC para modelo de demanda de servicios de salud al incluir variable de actividad física.

Ilustración No. 13

Curva ROC de regresión logística sobre demanda de servicios de salud, se incluye variable de actividad física



Fuente: ENSANUT 2011-2013

Elaboración: Diego Villacreses

Tabla No. 24

Matriz de Confusión de demanda de servicios de salud, se incluye variable de actividad física, punto de corte $\Pr(Y=1)$

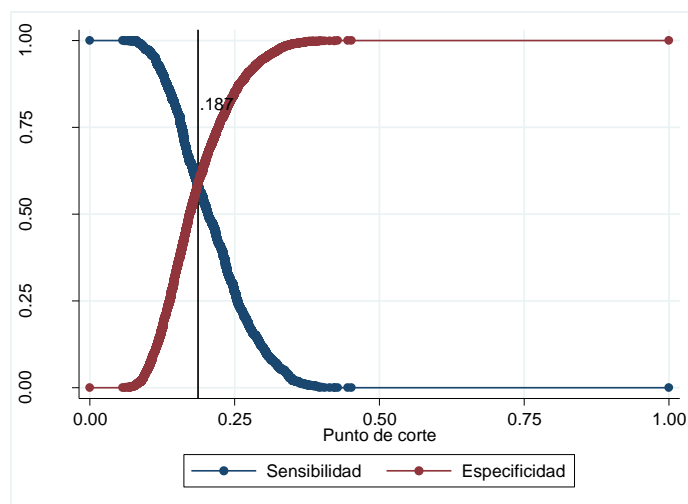
Clasificación con $\Pr(D) \geq .2114$	
Sensibilidad	47.94%
Especificidad	70.79%
Falso positivo	29.21%
Falso negativo	52.06%
Porcentaje de positivos incorrectamente clasificados	72.71%
Porcentaje de negativos incorrectamente clasificados	14.40%
Clasificación global	66.53%

Fuente: ENSANUT 2011-2013

Elaboración: Diego Villacreses

Ilustración No. 14

Sensibilidad y especificidad en relación a puntos de corte, demanda de servicios de salud, se incluye variable de actividad física.



Fuente: ENSANUT 2011-2013

Elaboración: Diego Villacreses

Tabla No. 25

Matriz de Confusión demanda de servicios de salud, se incluye variable de actividad física, punto de corte $\min(sens. - espec.)$

Clasificación con $Pr(D) \geq 0.1867$	
Sensibilidad	59.19%
Especificidad	60.48%
Falso positivo	39.52%
Falso negativo	40.81%
Porcentaje de positivos incorrectamente clasificados	74.49%
Porcentaje de negativos incorrectamente clasificados	13.37%
Clasificación global	60.24%

Fuente: ENSANUT 2011-2013

Elaboración: Diego Villacreses

Anexo 3

Sintaxis para la obtención de estimaciones

```
clear all
set more off
mata: mata clear
clear matrix
set mem 2g
set processors 8
set maxvar 30000

set seed 1234

*Programa para exportar a un .txt el "estat class"
cap program drop coso1
program define coso1
args arg1 arg2
qui{
tempvar bla1 bla2
gen str80 `bla1' = ""
gen str6 `bla2' = ""
local cp = `arg1'
replace `bla1' = "Clasificación con Pr(D)>= `round(`cp',0.0001)'" in 1
replace `bla1' = "Sensibilidad" in 2
replace `bla1' = "Especificidad" in 3
replace `bla1' = "" in 4
replace `bla1' = "Falso positivo" in 5
replace `bla1' = "Falso negativo" in 6
replace `bla1' = "Porcentaje de positivos incorrectamente clasificados" in 7
replace `bla1' = "Porcentaje de negativos incorrectamente clasificados" in 8
replace `bla1' = "" in 9
replace `bla1' = "Clasificación global" in 10
replace `bla2' = "" in 1
replace `bla2' = "`round(r(P_p1),0.01)'" in 2
replace `bla2' = "`round(r(P_n0),0.01)'" in 3
replace `bla2' = "" in 4
replace `bla2' = "`round(r(P_p0),0.01)'" in 5
replace `bla2' = "`round(r(P_n1),0.01)'" in 6
replace `bla2' = "`round(r(P_op),0.01)'" in 7
replace `bla2' = "`round(r(P_1n),0.01)'" in 8
replace `bla2' = "" in 9
replace `bla2' = "`round(r(P_corr),0.01)'" in 10
}
nois list `bla1' `bla2' if `bla1'!="
outsheet `bla1' `bla2' if `bla1'!=" using `arg2'.txt, replace noq nonames
end

*global dir "C:\Users\Deigo\Desktop\tesis"
global dir "C:\Users\pc\Desktop\Tesis\Bases de datos"

*Corrección sobre posible carater faltante en el directorio
if regexm("$dir", "[\]$")==0 global dir "${dir}\"
```

```

*Directorios de entrada y salida
global in "${dir}input"
global out "${dir}output"
if regexm("$in", "[\\$]")==0 global in "${in}"
if regexm("$out", "[\\$]")==0 global out "${out}"

global beta "ensanut_F1_informacion_general ensanut_F1_personas ensanut_F1_vivienda
ensanut_f6_fact_riesgo_adolescentes ensanut_f7_fact_riesgo_mayores ensanut_f8_actividad_fisica"
cd "${out}"

foreach w of global beta{
cd "${in}"
di in red "`w"
use `w', clear

if "`w" == "ensanut_F1_personas"{
    cap drop PD02
    foreach x of varlist pd* pse* pa*{
        local q_ =upper("`x")
        cap ren `x' `q_'
    }
}

*cap drop ci
gen ci = ciudad
drop ciudad
label var ci "Codigo a nivel cantonal"
format ci %06.0f
format zona %03.0f
format sector %03.0f
format vivienda %02.0f
format hogar %02.0f
cap format persona %02.0f
    global beta1 "ci zona sector vivienda hogar persona"
    foreach x of global beta1 {
        cap tostring `x', replace force usedisplayformat
    }
cap drop idhogar
egen idhogar = concat(ci zona sector vivienda hogar )
label var idhogar "Código identificador del hogar"

cap egen id_individuo = concat(ci zona sector vivienda hogar persona)
cap label var id_individuo "Código identificador de la persona"
cd "${out}"
qui save `w', replace
}

use ensanut_F1_personas, clear

collapse (mean) pw, by(idhogar)
tempfile factores_viv
save `factores_viv', replace

use ensanut_F1_vivienda, clear
merge m:1 idhogar using `factores_viv', nogen keep(3)

```

```
*****
*Calculo de quintiles económico*
*****
```

```
foreach x of varlist vi*{
  local a = upper("`x'")
  ren `x' `a'
}

qui recode VI02 (1/2=1) (nonmissing=0), gen(cv1)
qui recode VI03 (1/3=1) (nonmissing=0), gen(cv2)
qui recode VI04 (1/2=1) (nonmissing=0), gen(cv3)
qui recode VI05 (1/4=1) (nonmissing=0), gen(cv4)
qui recode VI06 (1/1=1) (nonmissing=0), gen(cv5)
qui recode VI07 (1/1=1) (nonmissing=0), gen(cv6)
qui recode VI08 (1/1=1) (nonmissing=0), gen(cv7)
qui recode VI10 (1/1=1) (nonmissing=0), gen(cv8)
gen cv9= VI12/ VI11
qui recode VI13 (1/1=1) (nonmissing=0), gen(cv10)
qui recode VI16 (1/1=1) (nonmissing=0), gen(cv11)
qui recode VI17 (1/1=1) (nonmissing=0), gen(cv12)
qui recode VI18 (1/1=1) (nonmissing=0), gen(cv13)
forvalues i=2001/2029 {
  qui recode VI`i' (1=1) (nonmissing=0), gen(cv`i')
}
factor cv* [aw = pw], pcf
predict proxy_index
xtile quint = proxy_index [aw = pw], nq(5)
mean cv*, over(quin)
*Elaboración: Freddy Llerena
```

```
tempfile viv
save `viv', replace
```

```
use ensanut_F1_personas, clear
merge m:1 idhogar using ensanut_F1_informacion_general, nogen keep(3)
merge m:1 idhogar using `viv', nogen keep(3)
```

```
*****
*Calculo la pobreza por NBI*
*****
```

```
clonevar vtv = VI02
clonevar v03 = VI04
clonevar v07 = VI06
clonevar v08 = VI07
clonevar v09 = VI08
clonevar h01 = VI11
replace h01 = . if h01 == 99
clonevar p24 = PD19B
clonevar p23 = PD19A
clonevar p21 = PD16
clonevar p02 = PD06
qui recode p02 (6= 5)(5=6)
clonevar p03 = PD03
```



```

clonevar p27 = PA01
clonevar v05 = VI05
generat pob_siise=1
*** Personas que viven en viviendas particulares
generat viv_parti=1 if (vtv>=1 & vtv<=8)
replace viv_parti=0 if (vtv>=9 & vtv<=17)
egen numper=count(pob_siise), by (idhogar)

/****1er componente: dependencia económica
Años de escolaridad del jefe de hogar (no hay valores de cero en el
grado más alto aprobado pero si valores 99 que no responden
según lo indicado por el me, cada año aprobado del centro de alfabetización
equivale a 2 años del antiguo sistema de educación*/
qui recode p24 (99=.)
generat escol1=0 if p23==1 | p23==3
replace escol1=p24*2 if p23==2
replace escol1=p24 if p23==4
replace escol1=p24+6 if p23==5
replace escol1=0 if p23==6 & p24==0
replace escol1=p24-1 if p23==6 & (p24>=1 & p24<.)
replace escol1=p24+9 if p23==7
replace escol1=p24+12 if p23==8 | p23==9
replace escol1=p24+17 if p23==10
replace escol1=escol1-1 if p21==1 & p23!=2 & (escol1>=1 & escol1<.)
replace escol1=escol1-2 if p21==1 & p23==2 & (escol1>=1 & escol1<.)
*Jefes de hogar con 2 años o menos de educación primaria
generate escje_ =1 if escol1<=2 & p02==1
replace escje_ =0 if (escol1>=3 & escol1<.) & p02==1
replace escje_ =0 if escje_==. & p02==1 & (p23==5 | (p23>=7 & p23<=10))
egen escje = max(escje_) if viv_parti==1, by(idhogar)

drop escje_
*Ocupados
generat ocup_=0 if p03>=10
replace ocup_=1 if (p27>=1 & p27<=5) & p03>=10
egen ocup = sum (ocup_) if viv_parti==1, by(idhogar) //Cantidad de ocupados en el hogar
drop ocup_
*Ocupados per cápita
generat ocupc=numper/ocup
*Más de tres ocupados por persona en el hogar u hogar sin ocupados
generat m3ocuxper=0 if ocupc<=3 & viv_parti==1
replace m3ocuxper=1 if (ocupc> 3 & ocupc<.) & viv_parti==1
replace m3ocuxper=1 if (ocup==0) & viv_parti==1
*Dependencia económica
generat depecA=0 if escje!=. & m3ocuxper!=.
replace depecA=1 if escje==1 & m3ocuxper==1
drop escje ocup ocupc m3ocuxper pob_siise

**** Segundo componente: hogares con niños que no asisten a clases
generat noasis_=0 if p21==1 & (p03>=6 & p03<=12)
replace noasis_=1 if p21==2 & (p03>=6 & p03<=12)
egen noasis = sum (noasis_) if viv_parti==1, by(idhogar)
cap drop hog_noasis
generat hog_noasis= 0 if noasis==0

```

```
replace hog_noasis= 1 if (noasis>= 1 & noasis<.)
drop noasis_ noasis
```

**** Tercer componente: materiales deficitarios

```
cap drop matdef
generat matdef=0 if (v03>=1 & v03<=5) & (v05>=1 & v05<=5)
replace matdef=1 if (v03>=6 & v03<=7) | (v05>=6 & v05<=7)
```

**** Cuarto componente: servicios deficitarios

```
cap drop agua_ade
generat agua_ade=0 if viv_parti==1
replace agua_ade=1 if viv_parti==1 & (v07==1 & v08==1) //Cuando se tiene agua de red pública y tubería dentro de la vivienda
cap drop serdef
generat serdef=0 if (agua_ade==1) & (v09>=1 & v09<=2)
replace serdef=1 if (agua_ade==0) | (v09>=3 & v09!=.)
drop agua_ade
```

**** Quinto componente: hacinamiento

```
generat perdor=numper/h01
replace perdor=numper if h01==0
cap drop hacina
generat hacina=0 if perdor<=3
replace hacina=1 if perdor>3 & perdor<.
drop perdor
```

**** Suma de las dimensiones

```
cap drop nbi
egen nbi=rsum(depec hog_noasis matdef serdef hacina)
replace nbi=. if nbi==0 & (depec==. | hog_noasis==.)
label var nbi "Necesidades Basicas Insatisfechas"
```

**** Cálculo de pobreza por NBI

```
cap drop pobre
generat pobre=0 if nbi==0
replace pobre=1 if (nbi>=1 & nbi<=5)
```

**** Cálculo de pobreza extrema por NBI

```
cap drop pobre_ext
generat pobre_ext=0 if (nbi==0 | nbi==1)
replace pobre_ext=1 if (nbi>=2 & nbi<=5)
```

**** Cálculo de pobreza extrema-no extrema por NBI

```
gen nbi1 = .
replace nbi1 = 1 if (nbi>=2 & nbi<=5)
replace nbi1 = 2 if (nbi>=1 & nbi<=5) & nbi1 == .
replace nbi1 = 3 if nbi == 0
cap drop idhogar
egen idhogar = concat(ci zona sector vivienda hogar )
ren nbi dimensiones_nbi
ren nbi1 nbi
label define nbi 1 "Pobre extremo" 2 "Pobre no extremo" 3 "No pobre", replace
label values nbi nbi
```

```

****Variables auxiliares
ren area a
recode a (2=0 "Rural") (1=1 "Urbano"), gen(area)
recode PD02 (1 = 1 "hombre") (2=0 "mujer"), gen(sexo)
drop if PD03 == .
ren PD03 edad
recode PD15 (1 = 1 "si") (2=0 "no"), gen(analf)
recode PD13 (1=1 "indígena") (2 3 4 = 2 "afro/negro/mulato") (5 = 3 "montubio") (6 8 = 4 "mestizo/otro") ///
(7 = 5 "blanco"), gen(etn)

recode ps02 (2=0)
cap drop prov
gen prov = real(substr(idhogar,1,2))
egen dominio = group( prov area )
gen upm = real(substr(idhogar,1,12))

svyset upm [pweight=pw], strata(dominio) vce(linearized)

recode VI14 (1 = 0) (else = 1), gen(agua_bb)

cap drop piso
**** Calidad de piso
recode VI05 (6 7 = 0) (nonmissing =1), gen(piso)
cap drop vtv
**** Tipo de vivienda
recode VI02 (1 2 = 1 ) (nonmissing = 0), gen(vtv)
**** Eliminacion de basura
recode VI10 (1 = 1) (nonmissing =0), gen(basura)
**** Cuarto exclusivo para cocinar?
recode VI13 (2 =0), gen(cocina)
**** Baño exclusivo
*recode h03 (2 3 = 0), gen(bano)

recode PSE01 (4 = 0) (1/3 = 1)
recode PSE07A (1/4 = 1) (nonmissing = 0)
gen seg = 1 if PSE01 == 1 | PSE07A==1
replace seg = 0 if seg == .
bys idhogar: egen seg_jh = max(seg)
bys idhogar: egen escol_jh = max(escol)
bys idhogar: egen etn_jh = max(etn)
bys idhogar: egen analf_jh = max(analf)

tab prov, gen(prov_)
tab etn_jh, gen(etn_jh_)

gen edad = .
forval x = 0 / 12{
local a0 = `x'*5
local a1 = `a0' + 4
label define edad `x' ""`a0'`a1'", modify
qui replace edad = `x' if edad >=`a0' & edad <=`a1'

```

```

}
replace edad = 13 if edad >64
label define edad 13 "65 en adelante", modify
label values edad edad

egen int_sex_gedad = group(gedad sexo)
egen int_sex_area = group(area sexo)
egen int_area_gedad = group(area edad)
egen int_area_agua_bb = group(area agua_bb)

tab int_sex_gedad, gen(int_sex_gedad_)
tab int_area_agua_bb, gen(int_area_agua_bb_)
tab nbi, gen(nbi_)
drop nbi_3
save test0, replace
*/

use test0, clear
      *enfermedad
recode ps55 (2=0)

gen enfermedad = 1 if ps02 == 1 | ps55 == 1
replace enfermedad = 0 if enfermedad == .

replace enfermedad = 0 if ps07 == 1 | ps07 == 2 | ps07 == 8
label var enfermedad "Enfermedad considerada suficientemente grave para atencion"

cap drop upm
egen upm = group(ci a zona sector)
svyset upm [pw = pw], strata(prov) vce(linear)
egen clus = group(upm prov)

svy: mean ps02, over(mes)
anova ps02 i.mes

*Días hospitalizado
      cap drop dh
gen dh = 0 if ps55 == 0
replace dh = ps58a*(365/12) + ps58b*(365/52.142857) + ps58c if dh ==.
      *imputación de días hospitalización por enfermedad persistente
replace ps66 = . if ps66 >30
replace dh = dh + ps66 if ps66 !=.
      *valores atípicos
qui summ dh if dh !=0, detail
replace dh = round(`= r(mean) + 2.25*(r(p75)-r(p25))') if dh >`= r(mean) + 2.25*(r(p75)-r(p25))' & dh !=.
dis `= r(mean) + 3*(r(p75)-r(p25))'
      *Visitas centro de salud
gen vs = 1 if ps06 == 1 | ps06 == 2
replace vs = 0 if vs == . & ps02 !=.
*replace vs =2 if ps36 ==1
      *Salud preventiva
gen sp = 1 if ps40 ==1
replace sp = 0 if ps40 != . & sp == .

```

```

*Generar Variables Dicotómicas para Demanda de Salud
foreach x of varlist dh vs sp{
  recode `x' (2/max=1), gen(d_`x')
}
gen canton = substr(ci, 1,4)
*merge m:1 canton using r2012, keep(3)

save test1, replace
cd "$out
*variables del hogar
*Imputacion de valores perdidos en variable de ingresos
use test1, clear
ren VIVIENDA viv

gen c = 1 if PA08 >=30000 & PA08 !=.
replace c = 0 if c == .
replace PA08 = . if c == 1 & edad<=19
bys idhogar: egen bla= max(c)
gen edad2 = edad^2
gen edad3 = edad^3
gen logingr = ln(PA08)
cap drop prov_*
foreach x of varlist etn prov quint{
  tab `x', g(`x'_)
}
reg logingr escol etn_2 etn_3 etn_4 etn_5 prov_2 prov_3 prov_4 prov_5 prov_6 prov_7 prov_8 prov_9 prov_10
prov_11 prov_12 prov_13 prov_14 prov_15 prov_16 prov_17 prov_18 prov_19 prov_20 prov_21 prov_22 prov_23
prov_24 quint_2 quint_3 quint_4 quint_5 edad edad2 area sexo agua cv2002 cv2004 cv2005 cv2006 cv2007 cv2009
cv2010 cv2011 cv2013 cv2016 cv2017 cv2020 cv2021 cv2022 cv2028 cv2029 cv1 cv2 cv3 cv5 cv6 cv7 cv8 if PA08
<30000 & edad >=20
predict coso
gen imput = 0 if c==1
foreach x of varlist escol etn_2 etn_3 etn_4 etn_5 prov_2 prov_3 prov_4 prov_5 prov_6 prov_7 prov_8 prov_9
prov_10 prov_11 prov_12 prov_13 prov_14 prov_15 prov_16 prov_17 prov_18 prov_19 prov_20 prov_21 prov_22
prov_23 prov_24 quint_2 quint_3 quint_4 quint_5 edad edad2 area sexo agua cv2002 cv2004 cv2005 cv2006 cv2007
cv2009 cv2010 cv2011 cv2013 cv2016 cv2017 cv2020 cv2021 cv2022 cv2028 cv2029 cv1 cv2 cv3 cv5 cv6 cv7 cv8{
  replace imput = imput+_b[`x']*`x' if c==1
}
replace imput = imput+ _b[_cons] if c==1
replace imput = exp(imput)
replace PA08 = imput if c==1
gen pers = 1

*****

* Años de Escolaridad del Jefe de Hogar *
*****

cap drop p24
cap drop p23
cap drop p21
cap drop p02
cap drop p03
cap drop p27

```

```

cap drop escol1

clonevar p24 = PD19B
clonevar p23 = PD19A
clonevar p21 = PD16
clonevar p02 = PD06
qui recode p02 (6= 5)(5=6)
clonevar p03 = edad
clonevar p27 = PA01

generat escol1=0 if p23==1 | p23==3
replace escol1=p24*2 if p23==2
replace escol1=p24 if p23==4
replace escol1=p24+6 if p23==5
replace escol1=0 if p23==6 & p24==0
replace escol1=p24-1 if p23==6 & (p24>=1 & p24<.)
replace escol1=p24+9 if p23==7
replace escol1=p24+12 if p23==8 | p23==9
replace escol1=p24+17 if p23==10
replace escol1=escol1-1 if p21==1 & p23!=2 & (escol1>=1 & escol1<.)
replace escol1=escol1-2 if p21==1 & p23==2 & (escol1>=1 & escol1<.)
ren escol1 escol_1
label var escol_1 "Schooling"

sort idhog
cap drop escojef
cap drop escola_jef

gen escojef = escol_1 if PD06==1
bys idhog: egen escola_jef = max(escojef)

gen esco_ = escol_1 if PD06==2
bys idhog: egen escola_cony = max(esco_)

collapse (sum) pers ingr_tot =PA08 (max) escola_jef escola_cony, by(idhogar)
gen ingr_pc = ingr_tot/pers
save ing, replace

cd "$out
use test1, clear

generat region=1 if (prov==1 | prov==2 | prov==3 | prov==4 | prov==5 | prov==6 | prov==10 | prov==11 | prov==17 |
prov==18 | prov==23)
replace region=2 if (prov==7 | prov==8 | prov==9 | prov==12 | prov==13 | prov==24 | prov==90)
replace region=3 if (prov==14 | prov==15 | prov==16 | prov==19 | prov==21 | prov==22)
replace region=4 if (prov==20)

drop if ps02 ==.

*Region
label var region "Region"
label define region 1 "Sierra" 2 "Costa" 3 "Amazonía" 4 "Insular" , replace
label values region region

```

```

cap drop a
decode region , gen(r)
decode area, gen(a)
tempvar r
egen `r' = concat(r a), punct(" - ")
encode `r', gen(region2)

cap drop _merg
merge m:1 idhogar using ing, nogen
recode vs (2 =1)
quantiles ingr_pc [iw= pw], n(10) gencatvar(q_ing_pc)
quantiles ingr_pc [iw= pw], n(10) gencatvar(q10_ing_pc)
quantiles ingr_pc [iw= pw], n(5) gencatvar(q5_ing_pc)

*Grupo etario (10)
drop if edad >= 98

gen edad10 = .
forval x = 1/10{
local a0 = `x'*10
local a1 = `a0'-9
dis "`a0' -- `a1'"
replace edad10 = `x' if edad>= `a1' & edad<= `a0'
label define edad10 `x' "`a0' - `a1'", modify
}
label values edad10 edad10
replace edad10 = 1 if edad10 ==0
label define edad10 1 "0 - 10", modify

*Enfermedad leve
recode ps07 (1 2 8 =1) (else = .), gen(leve)

*Enfermedades Crónicas
gen cron = 1 if ps03 == 12
replace cron = 0 if cron ==.

gen cron_l = 1 if ps03 == 12
replace cron_l = 0 if cron_l ==. | leve==1

*Logaritmo de ingresos
gen lnW = ln(ingr_tot)
*Cantidad de personas en el hogar
tempvar u
gen `u' = 1
bys idhogar: egen npers = sum(`u')
tempvar u
gen `u' = 1 if edad <=6
bys idhogar: egen npers06 = sum(`u')

tempvar u
gen `u' = 1 if edad <6
bys idhogar: egen npers05 = sum(`u')

```

```

tempvar u
gen `u' = 1 if edad >=6 & edad <=15
bys idhogar: egen npers615 = sum(`u')

tempvar u
gen `u' = 1 if edad >=65
bys idhogar: egen npers65 = sum(`u')

*Efectos no lineales edad
gen edad2= edad^2
gen edad3= edad^3

*Agua
recode VI06 (1 = 1) (nonmissing = 0), gen(agua)

*Servicios higienico
recode VI08 (1 = 1) (nonmissing = 0), gen(letrina)

*Piso
cap drop piso
recode VI05 (5 6 = 1) (nonmissing = 0), gen(piso)

*Cuarto exclusivo para cocinar
recode VI13 (1=1) (2=0), gen(cuarto)

*Se eliminan valores atípicos y NS/NR en ingresos
replace PA08 = . if PA08>30000

save base, replace

*****Empieza coso
use base, clear
drop quint
cap drop _*
sumdist proxy_index [aw = pw], ngp(5) qgp(quint)

***Grossman
    *Se corrige no informa día por media condicional a tipo de enfermedad
    recode p03 (0/20 = 1) (21/50 =2) (51/max =3), gen(gedad1)
    recode p03 (0/30 =1) (31/max =2), gen(gedad2)
    recode p03 (min/max =1) , gen(gedad3)

    levelsof ps03, local(test0)
foreach x of local test0{
    qui summ ps03 if ps03 == `x'
    if `=r(N)' >300 local disc gedad1
    if `=r(N)' <=300 & `=r(N)' >=90 local disc gedad2
    if `=r(N)' < 90 local disc gedad3
    levels `disc', local(test1)
    foreach x0 of local test1{
        dis "`x'-- `x0'"
        qui summ ps05a if ps03 == `x' & ps05a != 99 & `disc' == `x0'
    }
}

```



```

        replace ps05a = r(mean) if ps03 == `x' & ps05a == 99 & `disc' == `x0'
    }
}
gen D_enf = ps05b*7 + ps05a
gen _D_enf= -D_enf
replace D_enf = 0 if D_enf==.
drop if ps05b == 9
*recode D_enf (30/max = 30)
gen _lnTL = - ln(D_enf )
gen lnTL = ln(D_enf )
replace lnTL = lnTL+0.0001 if lnTL == 0

cap drop lnW_pc
gen lnW_pc = ln(ingr_pc)

cap drop lnW
gen lnW = ln( PA08)
gen lnD = ln(30 - D_enf)

*gen lnR
qui summ proxy_index
gen proxy_R = proxy_index + `=abs(r(min))'+0.001
gen ln_proxy_R = ln(proxy_R )

cap drop _lnTL2
gen _lnTL2 = - ln(D_enf +0.0001)

reg _lnTL2 escola_jef lnW_pc i.quint p03 sex [pw = pw], vce(cluster upm)
outreg2 using Grossman1.txt, replace
reg _lnTL2 escola_jef lnW_pc proxy_index p03 sex [pw = pw], vce(cluster upm)
outreg2 using Grossman1.txt, append
reg _lnTL2 escola_jef lnW_pc ln_proxy_R p03 sex [pw = pw], vce(cluster upm)
outreg2 using Grossman1.txt, append

reg _lnTL2 escola_jef lnW_pc proxy_index p03 sex [pw = pw] if p03>=18 & p03<=65, vce(cluster upm)
outreg2 using Grossman2.txt, replace
reg _lnTL2 escol lnW proxy_index p03 sex [pw = pw] if p03>=18 & p03<=65, vce(cluster upm)
outreg2 using Grossman2.txt, append
reg _lnTL2 escol lnW proxy_index p03 sex [pw = pw] if p03>=18 & p03<=65 & PSE01 == 1, vce(cluster upm)
outreg2 using Grossman2.txt, append
reg _lnTL2 escola_jef lnW_pc proxy_index p03 sex [pw = pw] if p03>=18 & p03<=65 & PSE01 == 1, vce(cluster upm)
outreg2 using Grossman2.txt, append

*Propuesta para controles
cap drop ps02_*
gen ps02_sin_emb=ps02
gen ps02_sin_enfmuj = ps02
gen ps02_sin_2m = ps02

replace ps02_sin_emb = 0 if ps03 == 8
replace ps02_sin_enfmuj = 0 if ps03 == 13

```

```

replace ps02_sin_2m = 0 if ps03 == 13 | ps03 == 8

foreach x of varlist ps02_sin_enfmuj ps02_sin_2m{
  replace `x' = 0 if ps03 == 14
}

mean ps02_sin_emb, over(sexo)
mean ps02_sin_enfmuj, over(sexo)
mean ps02_sin_2m, over(sexo)

foreach x of varlist ps02*{
  replace `x' = `x' *100
}

tabout quint sexo using t1_1.txt, replace c(mean ps02 ci) f(1c) svy sum
tabout quint sexo using t1_2.txt, replace c(mean ps02_sin_emb ci) f(1c) svy sum
tabout quint sexo using t1_3.txt, replace c(mean ps02_sin_enfmuj ci) f(1c) svy sum
tabout quint sexo using t1_4.txt, replace c(mean ps02_sin_2m ci) f(1c) svy sum

cap drop D_enf_sin_enfmuj
gen D_enf_sin_enfmuj = D_enf
replace D_enf_sin_enfmuj = 0 if ps02_sin_2m == 0

gen _lnTL3 = -ln(D_enf_sin_enfmuj+0.001)

reg _lnTL2 escol lnW proxy_index p03 sex if p03>=18 & p03<=65 [pw = pw], vce(cluster upm)
outreg2 using Grossman3.txt, replace
reg _lnTL3 escol lnW proxy_index p03 sex if p03>=18 & p03<=65 [pw = pw], vce(cluster upm)
outreg2 using Grossman3.txt, append

****Sección de demanda de la salud ampliado
replace ps02 = 1 if ps02 ==100
gen d_a = 1 if ps02 == 1 & ps06 !=5
replace d_a = 0 if d_a==.
gen d_a1 = 1 if ps02 == 1 & ps06 !=5 & ps06 !=3
replace d_a1 = 0 if d_a1==.

#delimit
global X "edad edad2 lnW proxy_index escola_jef i.etn area sexo
npers05 npers65 agua letrina piso cuarto seg";
#delimit cr

logit ps02 $X [pw = pw], vce(cluster upm)
outreg2 using sa_1.txt, replace
margins, dydx(*) post
outreg2 using sa_1.txt, append

logit d_a $X [pw = pw], vce(cluster upm)
outreg2 using sa_2.txt, replace
margins, dydx(*) post
outreg2 using sa_2.txt, append

logit d_a1 $X [pw = pw], vce(cluster upm)

```

```

outreg2 using sa_3.txt, replace
margins, dydx(*) post
outreg2 using sa_3.txt, append

```

```

*Enfermarse (1)
qui logit ps02 $X [iw = pw]
lroc, graphregion(fcolor(white)) ytitle(Sensibilidad) xtitle(1-Especificidad) msize(vsmall)
graph export sa_1.emf, replace
cap drop prob
cap drop sens
cap drop spec
cap drop diff_s
lsens, genprob(prob) gensens(sens) genspec(spec)
gen diff_s = abs(sens-spec)
sort diff_s
local pc1_1 = prob[1]
lsens, graphregion(fcolor(white)) ytitle(Correctamente clasificados) ///
xtitle(Punto de corte) legend(order(1 "Sensibilidad" 2 "Especificidad")) ///
xline(`pc1_1', lcolor(black) lwidth(medium) ) text( 0.82 `='pc1_1'+0.032' "'=round(`pc1_1',0.001)'" )
graph export ls_1.emf, replace
qui logit d_a $X [w = round(pw)]
qui summ d_a
local m = r(mean)
estat class, cutoff(`m')
coso1 `m' class_1_1
estat class, cutoff(`pc1_1')
coso1 `pc1_1' class_1_2

```

```

*Demandar salud (2)
logit d_a $X [iw = pw]
lroc, graphregion(fcolor(white)) ytitle(Sensibilidad) xtitle(1-Especificidad) msize(vsmall)
graph export sa_2.emf, replace
cap drop prob
cap drop sens
cap drop spec
cap drop diff_s
lsens, genprob(prob) gensens(sens) genspec(spec)
gen diff_s = abs(sens-spec)
sort diff_s
local pc1_2 = prob[1]
lsens, graphregion(fcolor(white)) ytitle(Correctamente clasificados) ///
xtitle(Punto de corte) legend(order(1 "Sensibilidad" 2 "Especificidad")) ///
xline(`pc1_2', lcolor(black) lwidth(medium) ) text( 0.82 `='pc1_2'+0.032' "'=round(`pc1_2',0.001)'" )
graph export ls_2.emf, replace
qui logit d_a $X [w = round(pw)]
qui summ d_a
local m = r(mean)
estat class, cutoff(`m')
coso1 `m' class_2_1
estat class, cutoff(`pc1_2')
coso1 `pc1_2' class_2_2

```

```

*Demandar servicios de salud (3)
qui logit d_a1 $X [iw = pw]

```

```

lroc, graphregion(fcolor(white)) ytitle(Sensibilidad) xtitle(1-Especificidad) msize(vsmall)
graph export sa_3.emf, replace
cap drop prob
cap drop sens
cap drop spec
cap drop diff_s
lsens, genprob(prob) gensens(sens) genspec(spec)
gen diff_s = abs(sens-spec)
sort diff_s
local pc1_3 = prob[1]
lsens, graphregion(fcolor(white)) ytitle(Correctamente clasificados) ///
xtitle(Punto de corte) legend(order(1 "Sensibilidad" 2 "Especificidad")) ///
xline(`pc1_3', lcolor(black) lwidth(medium) ) text( 0.82 `='pc1_3'+0.032' ``=round(`pc1_3',0.001)')
graph export ls_3.emf, replace
qui logit d_a1 $X [w = round(pw)]
qui summ d_a1
local m = r(mean)
estat class, cutoff(`m')
coso1 `m' class_3_1
estat class, cutoff(`pc1_3')
coso1 `pc1_3' class_3_2

```

*Modelos de Conteo

```

gen D_enf1 = D_enf
gen D_enf2 = D_enf
replace D_enf1 = 0 if d_a == 0
replace D_enf2 = 0 if d_a1 == 0

poisson D_enf $X [pw = pw], vce(cluster upm)
outreg2 using p_1.txt, replace
poisson D_enf1 $X [pw = pw], vce(cluster upm)
outreg2 using p_2.txt, replace
poisson D_enf2 $X [pw = pw], vce(cluster upm)
outreg2 using p_3.txt, replace
save base1, replace

```

*Se analizarán las regresiones anteriores en relación a actividad física***

```

use ensanut_f8_actividad_fisica, clear

```

```

ren pw pw_acti_fis
foreach x of varlist f8*{
qui replace `x' = . if `x'>60
}
foreach x of varlist f????{
qui recode `x' (2=0)
}
drop f8101
egen ac = rsum(f????)
global var0 "814 816 822 824 826"
global var1 "4 4 4 4.4 8"

```

```

tokenize "$var1"
foreach x of global var0{
    foreach y of varlist f`x`{*{
        qui if substr("`y'",8,1)=="a" replace `y'=`y'*60
    }
}
global dias "lu ma mi ju vi sa do"
local i 0
foreach x of global var0{
    local ++i
        foreach y of global dias{
            egen tiempo`x`y'= rsum(f`x`y'*)
        }
}
global var0 "814 816 822 824 826"

foreach x of varlist tiempo*{
    replace `x' = 0 if `x' <=10
}

foreach x of global var0{
    egen tiemp_`x' = rsum(tiempo`x'*)
}
local a ""
foreach x of global var0{
    if `x' == 826 egen acti_vigo = rsum(tiemp_`x')
    local a "`a' tiemp_`x'"
}
egen acti_moderada = rsum(`a')
*Jóvenes de 18 a 64
gen __001 = acti_moderada+ acti_vigo*2
recode __001 (0/149.9999=0) (150/299.999999999=1) (300/max=2), gen(cat_acti)

keep id_i cat_acti pw_
save acti, replace

use base1, clear
merge 1:1 id_individuo using acti

*Solo se añade actividad física
*Poisson
poisson D_enf $X i.cat_acti [pw = pw_acti_fis], vce(cluster upm)
outreg2 using p_1_1.txt, replace
poisson D_enf1 $X i.cat_acti [pw = pw_acti_fis], vce(cluster upm)
outreg2 using p_2_1.txt, replace
poisson D_enf2 $X i.cat_acti [pw = pw_acti_fis], vce(cluster upm)
outreg2 using p_3_1.txt, replace
*Logit
logit ps02 $X i.cat_acti [pw = pw_acti_fis], vce(cluster upm)
outreg2 using sa_1_1.txt, replace
margins, dydx(*) post
outreg2 using sa_1_1.txt, append

```

```

logit d_a $X i.cat_acti [pw = pw_acti_fis], vce(cluster upm)
outreg2 using sa_2_1.txt, replace
margins, dydx(*) post
outreg2 using sa_2_1.txt, append

logit d_a1 $X i.cat_acti [pw = pw_acti_fis], vce(cluster upm)
outreg2 using sa_3_1.txt, replace
margins, dydx(*) post
outreg2 using sa_3_1.txt, append

*Enfermarse (1)
qui logit ps02 $X i.cat_acti [iw = pw_acti_fis]
lroc, graphregion(fcolor(white)) ytitle(Sensibilidad) xtitle(1-Especificidad) msize(vsmall)
graph export sa_1_1.emf, replace
cap drop prob
cap drop sens
cap drop spec
cap drop diff_s
lsens, genprob(prob) gensens(sens) genspec(spec)
gen diff_s = abs(sens-spec)
sort diff_s
local pc1_1 = prob[1]
lsens, graphregion(fcolor(white)) ytitle(Correctamente clasificados) ///
xtitle(Punto de corte) legend(order(1 "Sensibilidad" 2 "Especificidad")) ///
xline(`pc1_1', lcolor(black) lwidth(medium) ) text( 0.82 `='pc1_1'+0.032' '='=round(`pc1_1',0.001)'"')
*graph export ls_1_1.emf, replace
qui logit d_a $X i.cat_acti [w = round(pw_acti_fis)]
qui summ d_a
local m =r(mean)
estat class, cutoff(`m')
coso1 `m' class_1_1_1
estat class, cutoff(`pc1_1')
coso1 `pc1_1' class_1_2_1

*Demandar salud (2)
logit d_a $X i.cat_acti [iw = pw_acti_fis]
lroc, graphregion(fcolor(white)) ytitle(Sensibilidad) xtitle(1-Especificidad) msize(vsmall)
graph export sa_2_1.emf, replace
cap drop prob
cap drop sens
cap drop spec
cap drop diff_s
lsens, genprob(prob) gensens(sens) genspec(spec)
gen diff_s = abs(sens-spec)
sort diff_s
local pc1_2 = prob[1]
lsens, graphregion(fcolor(white)) ytitle(Correctamente clasificados) ///
xtitle(Punto de corte) legend(order(1 "Sensibilidad" 2 "Especificidad")) ///
xline(`pc1_2', lcolor(black) lwidth(medium) ) text( 0.82 `='pc1_2'+0.032' '='=round(`pc1_2',0.001)'"')
graph export ls_2_1.emf, replace
qui logit d_a $X [w = round(pw_acti_fis)]
qui summ d_a
local m =r(mean)
estat class, cutoff(`m')

```

```

coso1 `m' class_2_1_1
estat class, cutoff(`pc1_2')
coso1 `pc1_2' class_2_2_1

*Demandar servicios de salud (3)
qui logit d_a1 $X i.cat_acti [iw = pw_acti_fis]
lroc, graphregion(fcolor(white)) ytitle(Sensibilidad) xtitle(1-Especificidad) msize(vsmall)
graph export sa_3_1.emf, replace
cap drop prob
cap drop sens
cap drop spec
cap drop diff_s
lsens, genprob(prob) gensens(sens) genspec(spec)
gen diff_s = abs(sens-spec)
sort diff_s
local pc1_3 = prob[1]
lsens, graphregion(fcolor(white)) ytitle(Correctamente clasificados) ///
xtitle(Punto de corte) legend(order(1 "Sensibilidad" 2 "Especificidad")) ///
xline(`pc1_3', lcolor(black) lwidth(medium) ) text( 0.82 `='pc1_3'+0.032' "="round(`pc1_3',0.001)'"')
graph export ls_3_1.emf, replace
qui logit d_a1 $X [w = round(pw_acti_fis)]
qui summ d_a1
local m =r(mean)
estat class, cutoff(`m')
coso1 `m' class_3_1_1
estat class, cutoff(`pc1_3')
coso1 `pc1_3' class_3_2_1

```

Elaboración: Diego Villacreses